Cross-Layer Platform for Dynamic, Energy-Efficient Optical Networks

Caroline P. Lai

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

Abstract

Cross-Layer Platform for Dynamic, Energy-Efficient Optical Networks

Caroline P. Lai

The design of the next-generation Internet infrastructure is driven by the need to sustain the massive growth in bandwidth demands. Novel, energy-efficient, optical networking technologies and architectures are required to effectively meet the stringent performance requirements with low cost and ultrahigh energy efficiencies. In this thesis, a cross-layer communications platform is proposed to enable greater intelligence and functionality on the physical layer. Providing the optical layer with advanced networking capabilities will facilitate the dynamic management and optimization of optical switching based on performance monitoring measurements and higher-layer attributes. The cross-layer platform aims to create a new framework for networks to incorporate packet-scale measurement subsystems and techniques for monitoring the health of the optical channel. This will allow for quality-of-service- and energy-aware routing schemes, as well as an enhanced awareness of the optical data signals.

This thesis first presents the design and development of an optical packet switching fabric. Leveraging a networking test-bed environment to validate networking hypotheses, advanced switching functionalities are demonstrated, including the support for quality-of-service based routing and packet multicasting. The investigated cross-layering is based on emerging optical technologies, enabling packet protection techniques and packet-rate switching fabric reconfiguration. Coupled with fast performance monitoring, the platform will achieve significant performance gains within the endeavor of all-optical switching. Allowing for a more intelligent, programmable optical layer aims to support greater flexibility with respect to bandwidth allocation and potentially a significant reduction in the network's energy consumption.

The ultimate deliverable of this work is a high-performance, cross-layer enabled optical network node. The experimental demonstration of an initial prototype creates a dynamic network element with distributed control plane management, featuring fast packet-rate optical switching capabilities and embedded physical-layer performance monitoring modules. The cross-layer box enables an intelligent traffic delivery system that can dynamically manipulate optical switching on a packet-granular scale. With the goal of achieving advanced multi-layer routing and control algorithms, the network node requires an intelligent co-optimization across all the layers.

The proposed cross-layer design should drive optical technologies and architectures in an innovative way, in order to fulfill the void between

the design of basic photonic devices and the networking protocols that use them. The performance of the entire network – from the optical components, to the routing algorithms and user applications – should be optimized in concert. This contribution to the area of cross-layer network design creates an adaptable optical pipe that is extremely flexible and intelligent aware of both the physical optical signals and higher-layer requirements. The impact of this work will be seen in the realization of dynamic, energy-efficient optical communication links in future networking infrastructures.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my research advisor, Professor Keren Bergman, for her sustained guidance, leadership, and support throughout my doctoral studies. This work would not have been possible without her trust and vision. Her encouragement and enthusiasm have been a source of inspiration and motivation during the last five years and will continue to drive my endeavors beyond this point.

I am grateful to Professor Tony Heinz, Dr. Jeffrey Kash, Dr. Daniel Kilper, and Professor Gil Zussman for graciously serving on my dissertation committee. Also, thank you to Professor Richard Osgood for serving on my original proposal committee.

I would like to express my thanks to several teachers and professors for the knowledge that I gained from them and for providing me various research opportunities in optics. They imparted upon me the drive to learn, and I owe my interest in the field of optics to them. For this, I must thank Professors Amir Helmy and Nazir Kherani of the University of Toronto, Professor Andrew Kirk of McGill University, and Professor Richard Osgood of Columbia University.

Undoubtedly, this dissertation would not have been possible without the work of the many talented researchers in the Lightwave Research Lab with whom I have worked closely. Thank you to Assaf Shacham, whose work and setup paved the way for a great deal of this research, for his valuable feedback, insight, and honest opinions on everything. I am grateful to Odile Liboiron-Ladouceur for her dedicated assistance and advice; her continued mentorship to this day is greatly appreciated.

Thanks to Benjamin Lee, for his leadership, all his help with experimental setups and measurements, and consistently being a role model of a great researcher. I have also had the pleasure of working closely with Daniel Brunina and Howard Wang, whose research has overlapped the most with mine and who have contributed greatly to this work: thanks for the helpful discussions and making the many hours spent in the lab fun. I am thankful to other Lightwave Lab members, including Aleksandr Biberman, Johnnie Chan, Ajay Garg, Gilbert Hendry, Noam Ophir, Kishore Padmaraju, and Michael Wang, for research discussions and their friendship during our shared years. To the newest class of lab members and to all the student researchers who will follow in the group: good luck with your research endeavors.

I must also acknowledge valuable, fruitful discussions with the postdoctoral fellows and visiting scholars who have overlapped with me in the Lightwave Lab. A special thank you to Franz Fidler for guiding me early on in research: he provided much of the foundation for this work and working with him was an absolute pleasure. I am grateful to Balagangadhar Bathula for providing much of the networking perspective for this work. I also appreciate the contributions of Cedric Ware, Lin Xu, and Wenjia Zhang in the past year.

I was very fortunate to have worked with many excellent researchers throughout the course of this work, including Ilia Baldine, Daniel Kilper, Peter Winzer, and Gil Zussman. They have all been enormously considerate with their time and willingness to share their extensive knowledge of their specific technical focus. Their input and feedback have unquestionably strengthened the research in this thesis. I must also thank the principal investigators of collaborating research groups associated with

CIAN, particularly Professors Yeshaiahu Fainman, Bahram Jalali, and Alan Willner, for their undeniable enthusiasm in pursuing mutually beneficial topics. Within all of these collaborations, I am grateful to the numerous students and researchers in these labs for interesting discussions and the effort they expended during the times we worked together.

I would also like to thank Jeffrey Kash and his exceptional research group at IBM T. J. Watson Research Center, including Fuad Doany, Daniel Kuchta, Benjamin Lee, Petar Pepeljugoski, Alexander Rylyakov, Laurent Schares, and Clint Schow. The team provided important feedback to this research, and my summer internship at IBM provided a valuable opportunity to learn more about optical link design within the realm of high-performance computing.

I also acknowledge fellowship support from the IEEE Photonics Society and equipment support from Polatis Inc. I would also like to thank the wonderful administrative staff who have been so helpful throughout my time at Columbia, namely Jill Forger, Chammali Josephs, Elsa Sanchez, and Azlyn Smith.

I would like to extend a heartfelt thank you to many good friends who have contributed to this work through their encouragement, words of advice, and support. Thanks for being my running partners, partaking in food adventures, and keeping me sane; thank you especially to those who help proofread this thesis.

Lastly, I am deeply indebted to my family, whose love and support have been unwavering throughout the years. To my family in Malaysia and England: thank you for your constant love. Thanks to my sister Stephanie for keeping me grounded. A very special thank you to my wonderful parents, to whom I dedicate this dissertation:

I am grateful for your endless guidance and love, and for the sacrifices that you have made so that I can be where I am today. I owe all of my successes to the both of you.

This thesis is dedicated to my parents.

Chapter 1

Introduction

Bandwidth-intensive applications and services are undoubtedly directly driving the requirement for the future Internet and networking infrastructures to support extremely high bandwidths and diverse data traffic flows. Researchers who created the first packet switching network (*i.e.* the Arpanet, sponsored by DARPA in 1969) initially envisioned their network as a small, distributed forum for academic research and communication, connecting a few thousand scientists and researchers [1]. Instead, their network quickly transformed to be the origins of today's Internet.

The subsequent approach to Internet design has been through Internet Protocol (IP) (layer 3) convergence, which has driven the development of the expansive Internet infrastructure that we know and use today. The strict isolation of the layers' functionalities allows researchers to optimize each layer's performance separately, enabling the rapid development and deployment of communication protocols and services. However, with the intense growth in data (driven mainly by high-bandwidth applications and other user traffic) and the exponential increase in the network's

energy consumption, it is becoming increasingly clear that this design model presents numerous limitations to sustaining future networking applications and end-to-end traffic. In order to sufficiently address these stringent and challenging performance requirements, it is necessary for the future network infrastructure to incorporate emerging physical-layer technologies and to diverge from traditional network design paradigms, specifically supporting a *cross-layer communications platform*.

1.1 Explosive Bandwidth Demands

It is unquestionably true that novel network designs will be needed to efficiently support the exploding demand for bandwidth-intensive applications and services. According to the Minnesota Internet Traffic Study (MINTS) [2] (Figure 1.1a) and other reputable bandwidth growth forecasts, the latest growth rate of North American Internet traffic as of the end-of-year 2008 hovers around 56% per year [3] (or about 2 dB per year [4]).

This growth trend extends beyond just the Internet to also complete network traffic (including cell carrier and telephone networks). The total traffic curve has the general shape as in Figure 1.1b [3]. Preceding the year 2000, voice dominated the network's traffic demand; however, since 2002, data has clearly overtaken voice as constituting the majority of network traffic (Figure 1.1b).

Recent trends indicate that networking equipment's bandwidth requirements are doubling every eighteen months [5]. Indeed, projections by both Cisco and the Discovery Institute maintain that the IP Internet traffic will grow internationally to a zettabyte (10^{21} bytes) by 2014, since the global IP traffic currently has an average growth rate of 20 exabytes (10^{18} bytes) per month [6]. This is reaching record levels

for both wireless and wired traffic. Figure 1.2 shows the projected forecasts in Internet traffic growth for consumers globally. Approximately 50% of the 20-exabyte/month growth in consumer traffic is composed of Internet video. This "exaflood" [7] of Internet and IP traffic, composed of video and rich media, necessitates a "deep transformation of the Internet capabilities and uses."

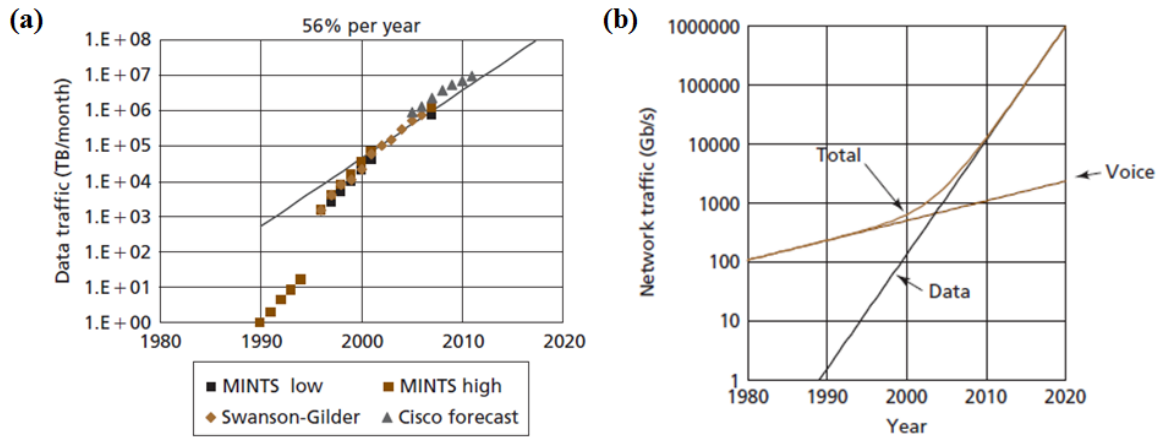

Figure 1.1: Bandwidth Growth Trends - (a) Bandwidth growth forecast as depicted by the Minnesota Internet Traffic Study; (b) Network traffic with respect to voice and data [3].

The trends in network traffic are due to the intense increase in the number of users, as well as the rapidly-changing ways users adopt the Internet in their daily lives. Bandwidth-intensive user-driven applications (*e.g.* online and/or on-demand gaming, data file transfer, high-definition (HD) video streaming, *etc.*) are increasingly challenging the bandwidths supported by today's networking infrastructure. Future data-centric and computing-oriented applications are also accelerating network traffic growth rates, especially with the rise in ubiquitous network/cloud computing. Furthermore, forward-looking, interactive, collaboration-focused applications such as

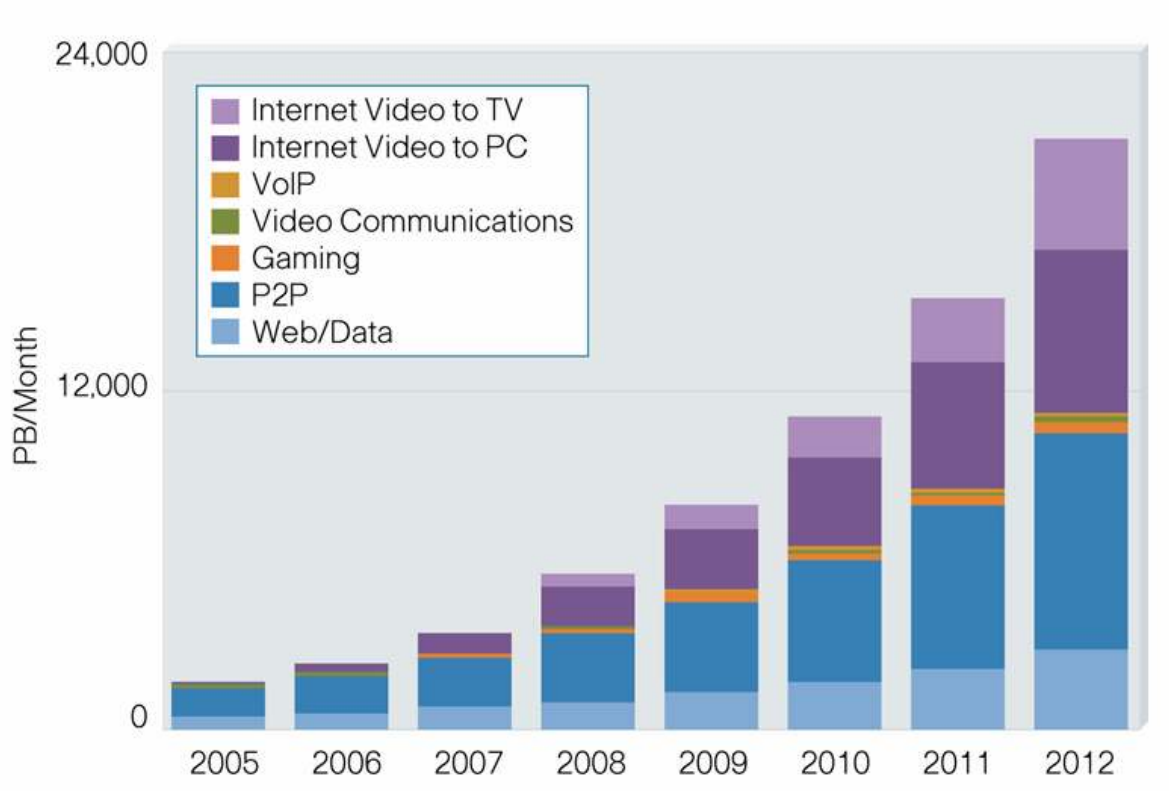

Figure 1.2: Consumer Internet Traffic Forecast - Cisco's forecasts for international consumer Internet traffic growth [6].

teleconferencing, telepresence, and telemedicine will require unprecedented broadband capabilities. The concept of telepresence [8] is expected to gain increasing importance as networks scale, aiming to deliver rich, immersive, multimedia, real-time applications with ultrahigh bandwidth demands. For example, Panasonic recently demonstrated its "Life Wall" product, allowing a television screen to occupy a large, wall-sized display (on the order of three 152-inch plasma displays). The product aims to provide a visual and immersive "window of information (and) communication tool" [9] for users to explore the Internet in large dimensions. The product – and other similar ones – will require high-speed access to an extremely large-capacity network.

Another real-time application, telemedicine, endeavors to provide interactive diagnostic and consultation services among leading physicians, medical professionals, and patients. Telemedicine will require the real-time transmission of HD (potentially uncompressed) video, high-resolution images, voice, and data in a very latency- and bandwidth-demanding scenario. Finally, the concept of three-dimensional (3D) television is also expected to be an important application that will gain extensive utilization. For instance, with the recent release of multiple 3D films and emerging holographic techniques that have attracted widespread attention [10], 3D display technologies with photorealistic qualities will begin to place extremely high-bandwidth demands on future networks. These multimedia services are representative of how typical high-bandwidth applications are driving the growth in the Internet.

To further emphasize this point: Sandvine, a network management company that studies Internet traffic patterns, emphasizes that we are seeing a dramatic shift from *asynchronous* applications to a greater demand for *real-time* applications. The majority

of bandwidth is dedicated to downloading files that are needed now (in real time), as opposed to files that are needed later (asynchronous). In the United States, the demand for Netflix exceeds YouTube, Hulu, and other peer-to-peer file-sharing protocols, accounting for almost 20% of downstream Internet traffic during peak home Internet usage hours [11]. Applications with the biggest jumps in traffic were dependent on real-time access such as streamed real-time audio and video (*i.e.* online gaming, Internet telephone programs, instant messages, *etc.*) These real-time applications will continue to find widespread utility and remain unquestionably the dominant drivers for data consumption on fixed and mobile networks worldwide. According to Cisco's forecasts [6], Internet video currently comprises greater than 43% of the total consumer Internet traffic – and still growing substantially. Further, nearly 64% of the world's mobile traffic will be video by 2013. The bandwidth requirements to efficiently enable these high-definition file transfers are orders of magnitude higher than other common Internet applications, potentially peaking around hundreds of Mb/s or even Gb/s – per user.

This shift in caliber of traffic demand will undoubtedly necessitate an Internet infrastructure (with the appropriate optical technologies) whose capabilities and functionalities far surpass those of today's network [12]. It is clear that current infrastructures cannot viably sustain this tremendous traffic growth in a scalable fashion [3, 13, 14, 15]. As asserted by Winzer [4], current electronically-multiplexed interface rates within routers are growing at approximately 12% per year, which cannot keep up with the sustained 56% of annual traffic growth. In order to truly sustain the bandwidth growth and mitigate these challenges, it may be required for routers to

adopt transparent, all-optical switching solutions.

1.2 Novel Photonic Technologies

As current networking systems scale, the major challenge lies in delivering this high-bandwidth, broadband user traffic with the necessary low communication latencies required by next-generation routers and network elements. Long-distance telecommunication systems have historically deployed lightwave solutions to extend the reach of communication links [16]. With the advent of low-loss fiber-optic technology in the 1970s [17], then with the invention of erbium doped fiber amplifiers (EDFAs) [18], incredible progress has been realized by leveraging the high bandwidth-distance product of optics. To meet the bandwidth demands and wide variety of applications, it is necessary for future networking designs to continue exploiting emerging physical-layer, photonic technologies [19]. By leveraging low-energy optical technologies, innovative architectures may be designed. The recent advances in optical technology enable us to redesign networks that will meet users' increasing demands.

Partially within the scope of a NSF-funded Engineering Research Center (ERC), the Center for Integrated Access Networks (CIAN) [20], and partially by other leading researchers in the industry (Winzer *et al.* , among others) [21, 22], there has been a tremendous push to design and develop the optical networks, technologies, subsystems, and devices that will be needed to address the demands of next-generation networks and systems. The optical technologies that will be necessary include innovative photonic devices (*e.g.* modulators, fast integrated switches, filters, lasers, *etc.*), as well as advanced optical switching technologies (*i.e.* hybrid optical/electronic switches), fast

performance monitoring solutions, optical routing algorithms, advanced modulation formats, and energy-efficient protocols. This work centers mainly on the latter networking-focused technologies.

With the drive of developing enhanced photonic switching technologies, optical packet switching (OPS) has been proposed as a promising, scalable, photonic approach for the construction of high-performance optical switching fabrics for future routers in the Internet [23], for both the access and carrier core domains. OPS fabrics can offer a programmable communications infrastructure for high-bandwidth, multiwavelength optical messages by allowing for the transparent transmission of broadband wavelength-striped optical messages with characteristically low latency and low power consumption [24, 25, 26]. The current transformative trend in deployed telecommunications networks is toward one of a purely packet-rooted architecture. Deploying optical packet switching fabrics in future network nodes will enable a flexible bandwidth-efficient data-centric Internet. This thesis investigates a possible OPS fabric architecture, presenting an experimental demonstration of the design in a networking test-bed environment and showcasing several advanced switching fabric functionalities that have been developed by the author.

1.3 Cross-Layer Paradigm

To further address the growing traffic demands along with the diverse variety of user bandwidth allocation requirements, the research community needs to look beyond the mere incorporation of innovative photonic technologies to additionally consider networking-based approaches wherein the network's physical-layer optical

switching can be optimized in concert with upper layers. The deployment of optical-domain based switching and transmission results in a reduction in the number of optical/electronic/optical (O/E/O) conversions. The system thus loses access to electronic regeneration techniques, which are key to maintaining adequate signal integrity. The overall network links are then more sensitive to physical-layer impairments.

Future networks can achieve reliable high-capacity connectivity without excessive overprovisioning using a cross-layer design paradigm [27, 28, 29, 30]. It is evident that addressing the immense user demand with a more intelligent, cross-layer enabled optical networking platform will also facilitate meeting the requirements for broadband quality-of-service (QoS) guaranteed user connectivity. This is particularly important as large-scale optical networks evolve towards a more packet-based infrastructure with a high level of required performance. Additionally, the cross-layer infrastructure will simultaneously take into account the "health" of the physical layer by catering to quality-of-transmission (QoT) operating conditions.

Networks will likely require a provision of services catering to the QoS of clients and end users. This will allow more efficient network resource allocation to be one of the essential ingredients to support the high-bandwidth traffic demands. By catering to the high data rates provided by a more flexible optical layer and to the QoS requirements as denoted by the IP layer in the proposed cross-layer fashion, we can envision the increased support of future multimedia and interactive applications.

1.4 Outline

The contributions and the innovative approach of this dissertation lie primarily in the development of a dynamic optical network node with a high level of flexibility and network functionality, as well as an increased physical-layer awareness of impairments, faults, and failures through the author's proposed cross-layer signaling infrastructure.

The remainder of this dissertation is organized as follows. The main objectives of this work are discussed in Chapter 2, wherein the key drivers and other related work are encapsulated. In Chapter 3, the optical switching fabric architecture that will comprise the envisioned optical node is described, as well as the general experimental implementation and complete test-bed environment in which the switching mechanisms and network hypotheses are validated. The increased network functionalities that were developed in this work are showcased in Chapter 4. Chapter 5 presents the advanced cross-layer communications platform, including experimental demonstrations and achieved switching mechanisms, with the encompassing performance monitoring work. The capstone of this dissertation is highlighted in Chapter 6, which discusses the design and implementation of a cross-layer enabled network node (the *CIAN Cross-Layer Box*) that features a fast reconfigurable optical switching fabric, advanced physical-layer functionalities, and performance measurement modules. First proposed and demonstrated by the author in this thesis, the initial prototype of the box supports the transmission of video using an Ethernet interface. This dissertation is summarized in Chapter 7 with an outline of ongoing and future work, in addition to a discussion of the contributions of this thesis. The ultimate endeavor of this work is to design and develop a highly-dynamic, QoS-aware optical cross-layer node that enables reliable and

robust high-capacity links for future data-centric networks.

Chapter 2

Objectives

In this chapter, an overview of the main objectives and key drivers of this work is provided. Other relevant research is explored and the focus is on how this thesis contributes to the huge body of work on optical technologies for future networks. Furthermore, this chapter specifically discusses the umbrella project of CIAN, as well as highlights the overarching goals of creating a dynamic physical layer and of enabling high network energy efficiencies.

2.1 Center for Integrated Access Networks (CIAN)

The scope of work covered by this dissertation is largely within the scope of CIAN [20] and its push to develop a novel optical solution for next-generation access/aggregation networks. Established in 2008, CIAN is an enormous multi-university research consortium with the mission of creating cost-effective transformative optical and optoelectronic technologies for optical access/aggregation networks that will provide high bandwidths throughout the Internet simultaneously with low cost and in an energy

efficient manner. This will allow data from any application at the edge, requiring any resource, at any time, to be seamlessly aggregated and interfaced with existing (and future) core networks cost-effectively.

2.1.1 CIAN's Vision

CIAN's vision is an intelligent Internet that fully meets the future demands for real-time, on-demand services by an increasing number of users, delivering information to every user at rates on the order of 10 Gb/s or higher at low cost and with high energy efficiencies [20]. This vision is propelled directly by the aforementioned accelerated growth in user broadband access amid the vast heterogeneity of applications, services, and emerging technologies. The ultimate goals of CIAN are twofold: (I) to develop architectures and systems that will meet the need of future data centers in terms of scalability, cost, and energy efficiency; and (II) to address the unprecedented requirements and capabilities of future networks by focusing on cross-layer intelligent aggregation networks, in order to optimize the network's energy efficiency and access/aggregation capabilities. The work in this thesis is largely involved in the latter goal (II), particularly in the "top-down" communications and networking thrust, which endeavors to drive the development and integration of optical components and devices to enable integrated subsystems. These modules will be co-optimized to cost-effectively provide high-data-rate services to heterogeneous (*i.e.* wireline optical and wireless) edge users. The key drivers are issues such as traffic aggregation, cross-layer optimization, and ubiquitous performance monitoring.

The goal of meeting the users' broadband needs will leverage the high bandwidths

and flexibility offered by fiber-optic communication systems, incorporating emerging optical technologies and network architectures to help meet industry's forward-thinking endeavors. The future design will demand advanced optical technology to attain a high level of dynamic tunability and programmability, and to offer the required high-speed broadband user connectivity while maintaining sufficiently low energy consumptions. The high-bandwidth, flexible, and agile networking functions and architectures will be realized using substantial bidirectional information exchange and optimization.

The innovative infrastructure will provide an integrated management and introspection tool to monitor the health of the optical data channels, in order to optimize overall end-to-end performance and to account for the network's energy consumption, as well the QoS requirements and QoT constraints of the future access/aggregation networks. The dynamic interaction between the physical-layer status, energy, and QoS will comprise a unique functionality that currently does not exist in today's networks.

The envisioned platform for bidirectional cross-layer information exchange will extract physical-layer monitoring measurements and provide these data to the higher layers which will use them for various functionalities. The goal is to achieve the following functional gains:

- state-of-the-art, energy-efficient devices and subsystems;
- energy-aware network architectures and routing capabilities;
- dynamic, real-time access to introspected data to allow for dynamic resource allocation;

- programmable flexibility accounting for application-specific QoS and optical QoT constraints;
- support for QoS in heterogeneous traffic environments (wireless and wired/optical);
- delivery of efficient, low-cost, high bandwidths to multiple users and applications (*i.e.* at the aggregation/core interface); and
- reliability and protection schemes via a cross-layer communication platform.

The mission of CIAN is to extend the high-data-rate handling capabilities of existing core networks further to the local access/aggregation networks to produce a low-cost technology that can reduce the energy consumption in the aggregation networks of the present and future Internet.

2.1.2 Evolution Toward a Mesh Topology

The growth in bandwidth demands is motivating industry to develop and evolve its network infrastructures to accommodate the unprecedented traffic flows. The CIAN vision is directly aligned with that of the current trend in industry: we are witnessing the evolution of the ring- and tree-based access/aggregation architecture to that of a more reliable mesh-type network topology [31]. Figure 2.1 depicts a high-level schematic of the current Internet architecture, using Synchronous Optical Networking/Synchronous Digital Hierarchy (SONET/SDH) rings and other tree architectures. In contrast, Figure 2.2 shows the focus of CIAN in developing a mesh-based design for the aggregation network. Mesh aggregation networks have been

studied for the mobile backhaul to enable an increase in physically disjoint network routes, thus leading to higher resilience within the network [32, 33]. This move to a mesh-centered network has been straightforwardly motivated by interactions with leading researchers and networking experts.

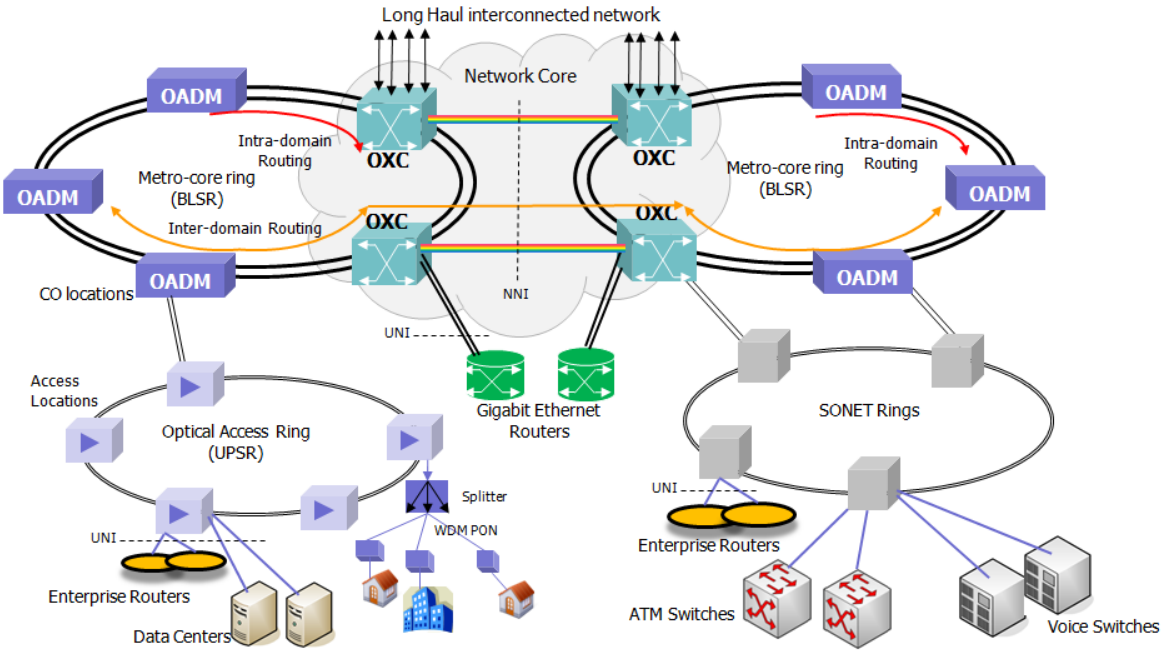

Figure 2.1: Current Network Architecture - Block diagram of the current network design.

2.1.3 Connection to the ERC Vision

The work described here leverages the devices and subsystems developed by other CIAN-affiliated institutions, as required by the top-down drivers. For example, CIAN-developed optical performance monitoring (OPM) subsystems (*e.g.* [34] used in this work [35]) allow for the introspective access to the optical layer and yield new network

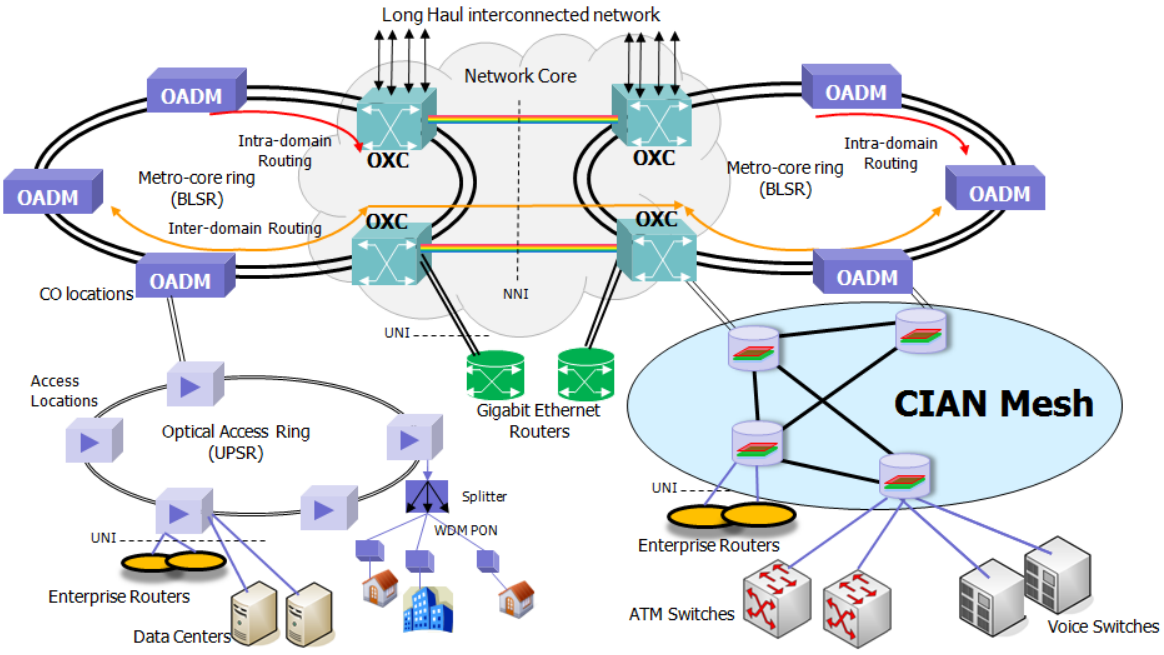

Figure 2.2: Mesh-Based Aggregation Network Architecture - Block diagram showing how the aggregation network is moving toward a mesh-centered design.

optimization possibilities [36]. OPM is an advantageous technology that can be exploited for advanced network functionality; seminal work has been performed on the subject of OPM by Kilper *et al.* [37]. The proposed introspective technologies can detect physical-layer degradations in real-time and feedback the performance information to higher layers to help ensure network reliability, which may particularly important as optical data rates continue to scale to meet the high-bandwidth traffic demand.

This work fits into the overall research agenda of the ERC by specifically addressing the development of a bidirectional information exchange that optimizes the network's energy efficiency, while supporting both QoS classes from the IP layer and QoT guarantees from real-time optical-layer introspection. The principal idea is to jointly optimize the data's QoS and energy metrics in terms of network performance.

The author's key contribution to this research effort is the development of a cross-layer optimized test-bed environment, creating a seamless, high-bandwidth, intelligent, programmable optical access/aggregation network node (Figure 2.3) composed of CIAN *cross-layer boxes* (CLBs), also known as *CIAN Boxes*. The CIAN Box is based on the flexible optical packet switching fabric platform described in Chapter 3, and will allow for an optical aggregation node that is capable of real-time monitoring and on-the-fly reconfiguration, incorporating various fast (optical) performance monitoring solutions (discussed in Chapter 5). A high-level vision for the CLB is given in Figure 2.4. The box will be fully controlled using an optical control and management plane for routing optimization. The goals of the CIAN Box are to achieve low energy consumption, fast nanosecond scale optical switching, and increased efficiency in bandwidth utilization

using QoS provisioning. The programmability of the switching fabric will be leveraged in the future to demonstrate energy-efficient routing algorithms and/or architectures designed through various network modeling and simulation projects.

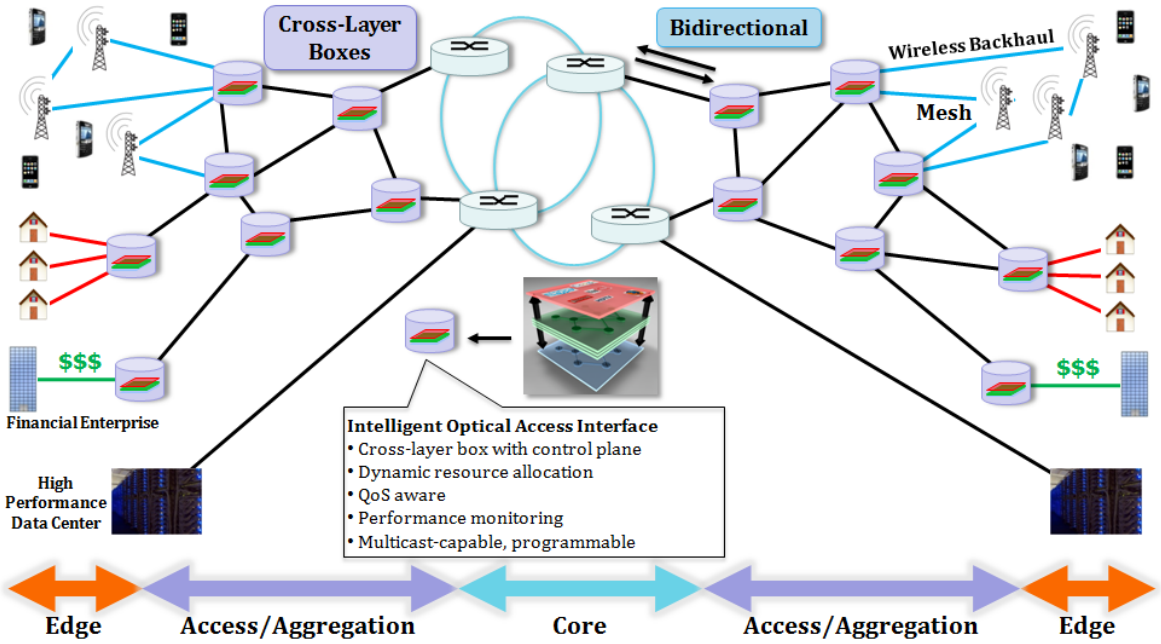

Figure 2.3: Cross-Layer Network Architecture - Schematic of envisioned network, including cross-layer boxes.

This CLB node serves as an intelligent transparent interface between the high-bandwidth core and heterogeneous (wireless/wireline) edge nodes, yielding a modular and programmable platform enabling cross-layer networking capabilities. The node is envisioned to have the ability to route (and possible optically multicast) wavelength-striped optical data traffic in a reconfigurable fashion, while simultaneously accounting for the physical-layer performance and the energy consumption of the inline optical and electrical components. The node will scale in the future to interconnect heterogeneous

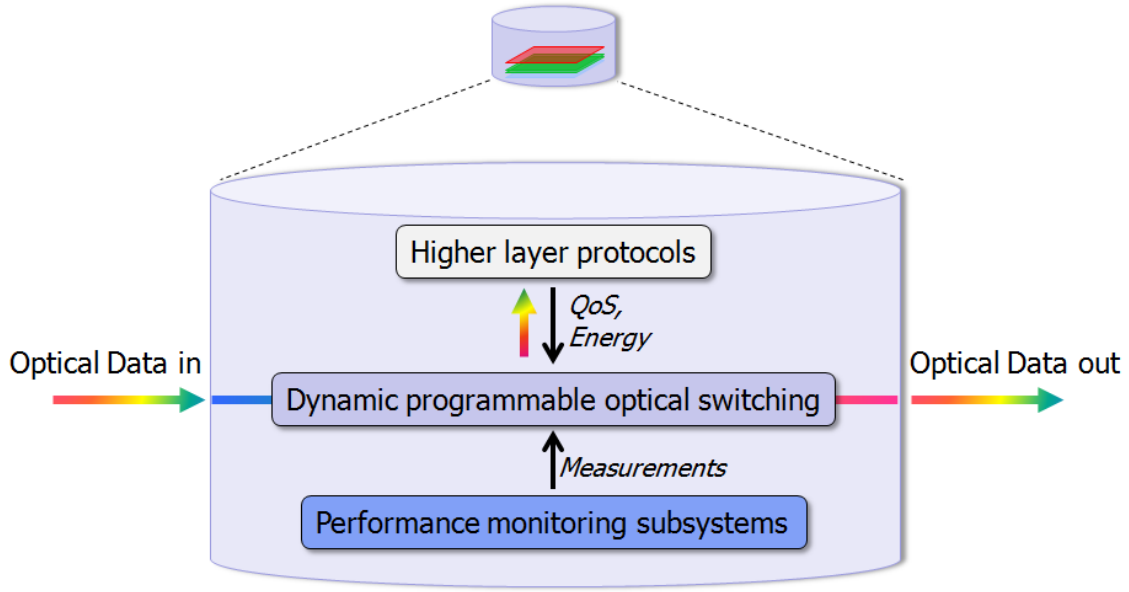

Figure 2.4: Cross-Layer Box - Basic block diagram of a single cross-layer box.

edge nodes, transparently aggregating wireline and wireless users at different bit rates, modulation formats, and priority requirements. One of the major challenges is the integration of programmable capabilities that can directly leverage knowledge of the optical signals in a cross-layer way. The physical-layer performance will be evaluated using CIAN-developed optical performance monitors that can measure parameters such as the bit-error rate (BER), optical-signal-to-noise ratio (OSNR), chromatic dispersion (CD), polarization-mode dispersion (PMD), *etc.* These signals can create a feedback path to higher layers to allow for packet protection, rerouting, or correction; ultimately, the cross-layer signaling will allow for traffic and network routing optimization.

The design and first demonstration of a CIAN Box is discussed in greater detail in Chapter 6. The initial design of the box at CIAN's cross-layer test-bed (at Columbia) includes high-capacity networking and programmable cross-layer functionality. It

features an optical packet multicast-capable switching fabric that truly exploits the high level of photonic capabilities enabled by innovative CIAN-driven optical technologies.

2.2 Intelligent Dynamic Physical Layer: Related Work

Many view the future Internet as an extremely dynamic networking environment that will provide incredibly high-speed access and connectivity, supporting a plethora of high-bandwidth services and applications [38]. The next-generation infrastructure should be flexible to the needs of users, transparent, and highly efficient in terms of energy consumption. The design of future networking functionalities and capabilities will be executed in accordance with users' requirements for high-bandwidth applications, *i.e.* the underlying optical network should be able to dynamically optimize its performance based on the requirements from the higher network layers. Researchers are aiming to bring an inherent dynamic capability to the physical layer through various switching techniques, software frameworks, and impairment-aware network planning tools.

To this end, there has been an incredible body of work devoted to addressing the dynamicism of the optical layer. The research community has long analyzed the desired characteristics and features of the future Internet (*i.e.* in a "clean-slate" redesign of the Internet) and explored ways in which they may viably conceive realizing the necessary architectures and optical technologies. In general, the community is in agreement that by migrating higher-layer functionalities to lower layers in the Open Systems Interconnection (OSI) protocol stack, there is the possibility to achieve greater

provisioning of optical resources with reduced cost (specifically according to Verizon in [39]). The inverted pyramid in Figure 2.5 shows that networks can achieve the lowest cost-per-bit in the case that data is switched on lowest (optical) layer. Thus, the consensus is that traffic should be managed and switched at the lowest possible OSI layer, further motivating all-optical switching for future networks.

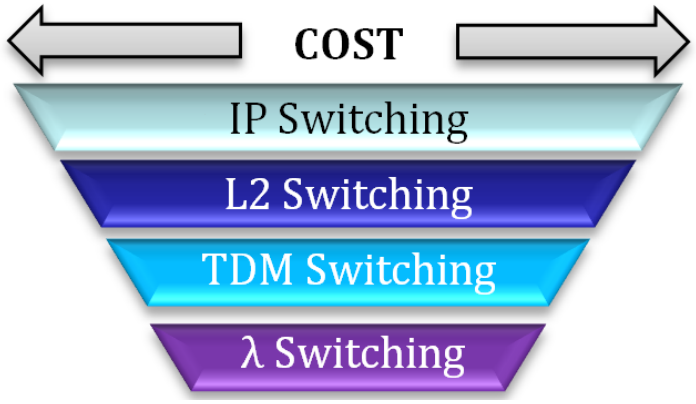

Figure 2.5: Switching Cost Pyramid - Block schematic of relative switching costs [39].

Multi-layer traffic engineering is an area of current research that is gaining attention. This uses the premise that the faster the switching can occur (potentially all-optically), the better the overall network performance [40]. Indeed, by switching at the lowest possible layer of the OSI stack, significant reductions in energy consumptions may be achieved. The typical energy-per-bit values of a few switching devices (deployed at different OSI layers) are given in Table 2.1 [41, 42]. One can observe that the cost-per-bit is greatly reduced by exploiting switching at the lower layers. Enabling multi-layer traffic engineering, in conjunction with integrated, packet-level QoS control, can then facilitate the design of a high-performance, multi-layer, multi-granular optical

Table 2.1: The energy consumption of several switching devices [41].

Switch Technology	Energy-per-bit
IP Router	10 nJ/bit
TDM Switch	1 nJ/bit
WDM Switch	0.5 nJ/bit

router [43]. This implies that the network requires an intelligent balance between all layers, achieving an enhanced multi-layer (cross-layer) coordination to provide reduced cost.

2.2.1 Optical Packet Switching

From the perspective of physical-layer switching, optical burst switching (OBS) comprises a feasible alternative for creating fast bursts of data that are switched all-optically using electronic control-plane processing, which may be hard to implement [23]. Optical flow switching (OFS) can also act as a means to realize end-to-end long-duration lightpaths between users, though requiring complex scheduling algorithms [44, 45, 46]. Dynamic optical circuit switching (DOCS) is yet another proposed solution to the switching problem, which yields high-bandwidth pipes between edge users and the backbone network, and features bandwidth-on-demand functionalities [47]. This results in more flexibility for scheduling routing requests and for adjusting to the required data rates for specific applications. These capabilities are obviously needed for future networks and is envisioned for this proposed approach. The optical switching technology discussed here in this work is optical packet switching (OPS), which can dynamically process, forward, and route packets directly on the

optical layer, leveraging wavelength-division multiplexing (WDM) to achieve high aggregate bandwidths [48]. OPS can straightforwardly address the optical packet transport evolution that is currently disrupting today's transport networks [49], and that is gaining increasing attention by key service providers and the optical networking industry (such as Light Reading, among others). The drive from industry to sufficiently transport data packets directly on the physical layer strengthens the viability of OPS.

2.2.2 Software Initiatives

The Global Environment for Network Innovations (GENI) [50] initiative is also examining software solutions through a clean-slate redesign of the networking infrastructure. From the perspective of these software innovations, the Services Integration, controL, and Optimization (SILO) architecture has been proposed as a cross-layer enabled design to address the notion of rigid network layering [27, 51]. SILO is a software framework that allows applications to create *silos* from application-specific building blocks, featuring gauges, knobs, and other tuning algorithms to cater to physical-layer performance. Under the GENI initiative [50], SILO has been integrated with Breakable Experiment Network (BEN), a metro-scale optical network test-bed in North Carolina. With the help of the author, an optical-power-aware video streaming experiment was performed, allowing for power fluctuations within a link supporting the transmission of video data to be compensated by an optical amplifier [52]. The push for dynamic restoration capabilities of the physical layer is evident from the SILO framework research effort.

2.2.3 Impairment-Aware Routing

From the perspective of network routing, there has been extensive simulation and experimental work on designing physical-layer impairment-aware routing algorithms by leading networking researchers [53, 54, 55, 56, 57]. As a case study example: simulations that are a part of the European Dynamic Impairment Constraint Networking for Transparent Mesh Networks (DICONET) project explore algorithms incorporating physical impairments in network planning through the measurement of signals' QoT via the BER [28, 58, 59]. For transparent networks, the lack of regeneration technologies results in physical impairments severely degrading the lightpaths' QoT. Cross-layer techniques are thus required for these transparent networks to incorporate physical-layer considerations. The QoT is estimated offline using the Q-Tool such that the network routing algorithms can assign Q-factor costs to links [58]. Q-Tool accounts for both static and dynamic impairments; static metrics include amplified spontaneous emission (ASE) and PMD. Dynamic impairments are dependent on other preexisting lightpaths, including crosstalk and nonlinear effects. The same researchers have also shown impressive experimental results on a 14-node network test-bed implementing their impairment-aware control plane techniques using integrated real-time QoT estimators [60], where they show that the bottleneck of QoT processing lies in the field-programmable gate array (FPGA). An enhanced version of this hardware platform has provided significant performance improvements [61]. The work performed by DICONET is laudable; it shows interesting and relevant progress to enabling cross-layer routing schemes that can incorporate physical-layer transmission factors. In the future, the author anticipates that CIAN will also demonstrate such advanced

network functionalities with dynamic physical-layer monitoring, while simultaneously accounting for application-specific QoS constraints in a cost-effective way for achieving high capacities in access/aggregation networks.

2.2.4 Advanced Modulation Formats

As a final point of this research survey, the author cannot fail to note the recent progress in coherent transmission and detection. While advanced modulation techniques is not directly addressed in this thesis, a great deal of positive and encouraging work has been performed in this area for telecommunication networks. To sufficiently carry the huge bandwidths that will be required, high-speed interfaces will be necessary; electronically-multiplexed interfaces have transitioned from direct to coherent detection [62]. To enable a truly dynamic physical layer, performance monitoring devices must not only measure various OPM metrics, as well as BER (similar to the DICONET project), but also leverage advanced digital signal processing with potentially coherent techniques [63, 64]. This relevant area of research is not covered here; however, the author believes that future work within CIAN and planned activities for the implemented cross-layer test-bed should address this issue.

2.2.5 Cross-Layer Communications

With the goal of realizing a dynamic physical layer, the cross-layer platform is presented by the author in this work as a seamless way to allow higher-layer applications and network routing algorithms to be complementary with – and concurrently optimized with – physical-layer characteristics and performance awareness. This will enable a

highly efficient, dynamic, more intelligent networking solution for next-generation IP networks and network routing applications. The author envisions a cross-layer design as shown in Figure 2.6, where the programmable optical layer can dynamically interact with higher network layers, creating a bidirectional information flow. Differentiated QoS requirements can flow downwards to the physical layer such that the data's QoS class can directly impact packet handling in the optical layer. Correspondingly, packet- or flow-level performance monitoring (PM) (or OPM) devices can extract real-time physical-layer performance, possibility indicating isolated signal impairments, and send these measurements upward in the network stack for higher layers to act upon. The architecture truly leverages physical-layer PM measurements (*e.g.* via extraction of the BER, OSNR, *etc.*), indicating possible signal degradations. The ultimate platform will endeavor to incorporate, drive, and exploit the emerging revolutionary and heterogeneous advances in physical-layer technologies, by allowing the integration of novel, flexible optical devices and monitoring subsystems directly in the physical layer to provide substantial performance gains for the overall network. The network will be able to holistically and intelligently recover from failures, manipulate optical data based on the signals' performance, and efficiently allot bandwidth.

2.3 Quality-of-Service Support

The cross-layer design will be used for routing, bandwidth allocation, and flow control, and this platform provides maximum benefits when exploiting a dynamic, programmable optical layer. To overcome the limitations currently stemming from rigid network layering, the redesign must consider the current notion that each network

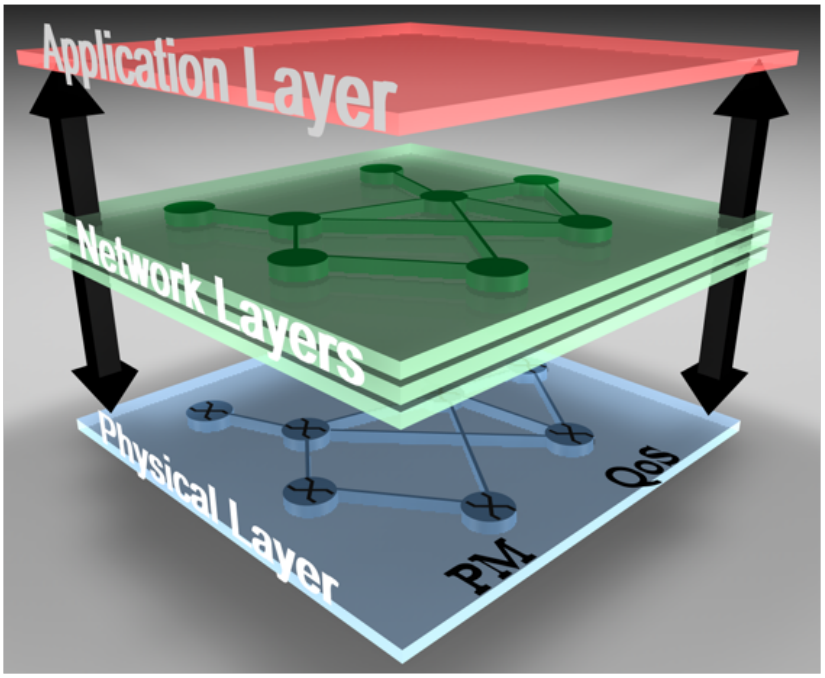

Figure 2.6: Cross-Layer Protocol Stack - Envisioned cross-layer-optimized network stack, supporting a bidirectional signal exchange between the network layers; the QoS-aware physical layer uses integrated performance monitoring devices.

layer is designed to provide bare-bones functionality that is used by the layers above it while essentially hiding information from the lower layers. Each layer is essentially blind to the layer below while providing a service to the layer above. For instance, the network layer currently provides only a *best-effort* routing, making no guarantees on packet delivery. These guarantees are implemented within the transport layer, which in turn views the network layer as a best-effort medium. By enabling the optical layer and the underlying optical devices to be more aware of the network parameters used by the routing layers, a more dynamic physical layer can then be achieved that can support more elastic and flexible QoS assurances.

For example, the state-of-the-art optical networking technologies that are being developed concurrently to this network redesign effort can provide substantial functionalities to meet numerous QoS requirements. Allowing the broadband applications to account for the QoS may provide many advantages [65]. However, when using the current IP in the layered design, applications have no straightforward means of deriving performance benefits through QoS capabilities. Conversely, network layers view the optical substrate as a "black box" and are forced to handle all packets in identical fashion, irrespective of their applications' transport needs (*i.e.* latency, throughput, resilience, *etc.*) This work develops an innovative integrated instrumental tool that uses physical-layer performance monitoring measurements as a basis for bidirectional information exchange and traffic engineering, to achieve programmable flexibility on application-specific QoS requirements and physical-layer aware network routing capabilities.

The exchange of management and control information between the physical data

layer and the above routing layers allows the network to adapt itself on-the-fly to the needs of higher-layer applications with different QoS requirements, with the added ability to adapt dynamically and seamlessly to changing networking conditions. The platform can be used to reconfigure packet routing based on performance monitoring measurements (*i.e.* QoT) such as BER degradations in the optical domain, packet loss, link failure, and other complementary OPM metrics. Dynamic varying QoS class requirements can then be offered while accounting for these performance degradations to yield more dynamic behavior on the optical layer. In this way, IP-caliber best-effort and high-priority classes can be supported directly on the optical layer.

An important note should also be made at this point regarding the distinction between QoS and QoT. This work is both QoS-aware (*i.e.* referring to the differentiated traffic classes and to network-level quality issues) and QoT-aware (*i.e.* referring to physical impairments and to the quality of transmission at the physical layer). The QoS requirements envisioned in this work are directly encoded by the QoS classes supported by the higher layers (such as priority, reliability, *etc.*) This facilitates application-specific information to flow downward in the OSI stack to affect optical-layer routing; accounting for this metric aims to reduce the number of retransmission required by higher layers in the stack (*i.e.* by the transport layer). This should be clearly differentiated from advanced QoT requirements, which are extracted from optical measurements as an indicator of physical-layer performance. The QoT, which may be related to the signals' BER, OSNR, PMD, or other OPM metrics, is not directly related to the QoS, unlike in other work. For example, [66] discusses the design of optical QoS for OBS networks and [67] studies the inclusion of BER and latency

thresholds in wavelength routing algorithms. A QoS-aware scheme supporting dynamic bandwidth allocation has also been investigated for wireless and passive optical access networks [68]. These external references to optical QoS parameters are actually interpreted here as designs supporting QoT requirements. Some related research in DICONET adopts the similar terminology as in this work [69]. An integrated framework for differentiated QoS control aims to create improved network resource usage and allocation, allowing for better traffic quality to be support on the optical layer with advanced coordination across all the network layers [43].

The goal here is to co-optimize optical packet routing based on both the QoS from the higher network layers, as well as the measured QoT which may be extracted on a packet-by-packet basis from the optical data.

2.4 Energy: Unsustainable Growth

A powerful driver for this work lies in addressing the power consumption of today's networks. The energy bottleneck has long been a key driving force for developing optical interconnects in high-performance computing systems, and is now becoming a limiting factor in telecommunication networks. With the increasing number of Information and Communication Technologies (ICT), the energy consumption associated with telecommunication networks is predicted to grow exponentially in the next decade [70, 71, 72].

In 2007, the carbon footprint of worldwide ICT was stated to be approximately 2% and expected to grow to 4% by the year 2020 [73]. Figure 2.7 depicts an energy consumption forecast for telecommunication networks [72]. This is mainly driven by the

widespread ubiquity in broadband applications as well as the growth in mobile devices. The networking community (including the GreenTouch Consortium [74], among others) is now dedicating a tremendous research effort to addressing the energy efficiency of future telecommunication networks: namely, how to sufficiently and effectively support the high-bandwidth services required by future applications while maintaining minimal energy consumptions [75, 76].

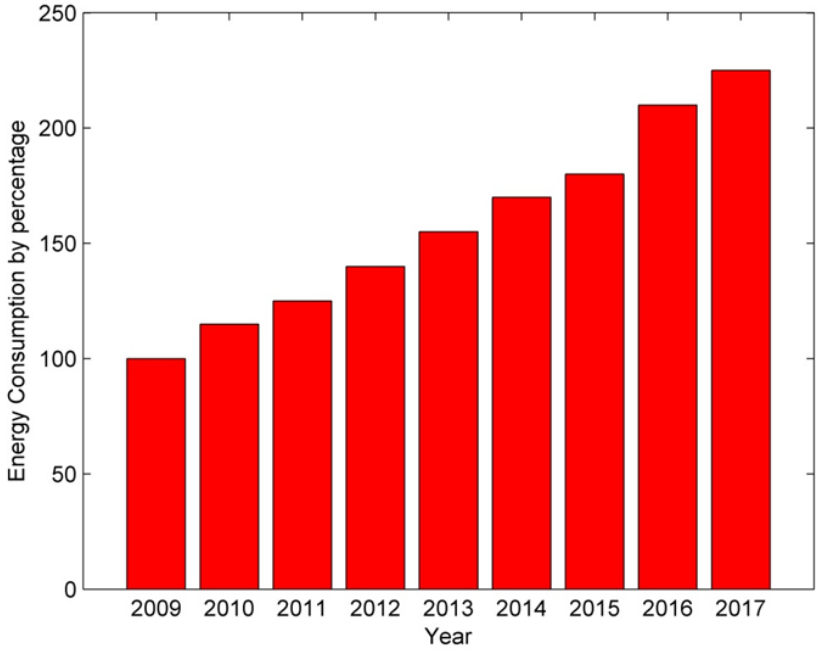

Figure 2.7: Telecommunication Energy Trends - Prediction of the energy consumption growth (by %) of telecommunication networks [72].

With the increasing percentage of total energy consumed by contemporary network equipment [77, 78, 79] (Table 2.2), it is evident that minimizing the total power consumption is a major driver for future networking infrastructures. The following figures have been presented by collaborators within GreenTouch (specifically

Kilper [40]). Figure 2.8 depicts the baseline business-as-usual forecasts for the estimated power-per-user for various parts of the network, including the performance of state-of-the-art devices and a 10% per year improvement yielded from Moore's Law. Even with the base-case improvements applied uniformly to 2017 (Figure 2.9), the network efficiency yields a flat power-per-user trend for the next decade.

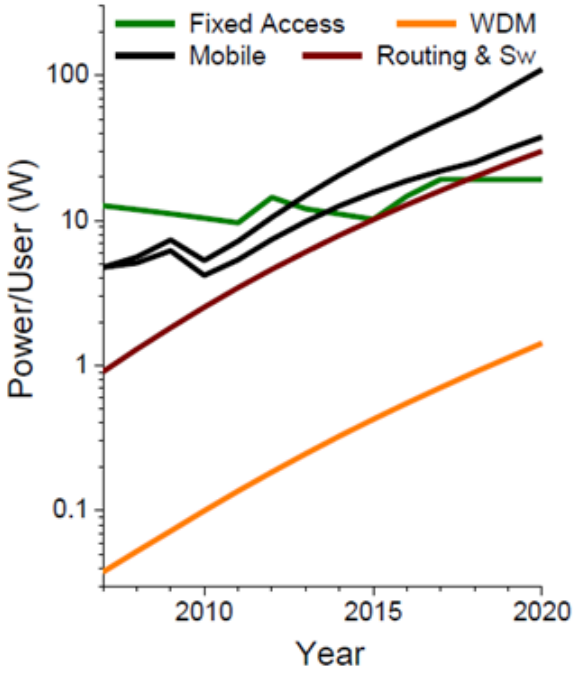

Figure 2.8: Business-As-Usual Efficiency Trends - Energy forecasts assuming no significant means for efficiency improvements [40].

Given the massive bandwidth growth and the accelerating energy consumption, following current incremental business-as-usual network technology advances cannot continue to meet video-driven bandwidth demands at a sustainable energy and cost [42]. The bottom line is that the networking community must adopt radically new energy-efficient, low-cost networking technology solutions to sustain explosive growth in user

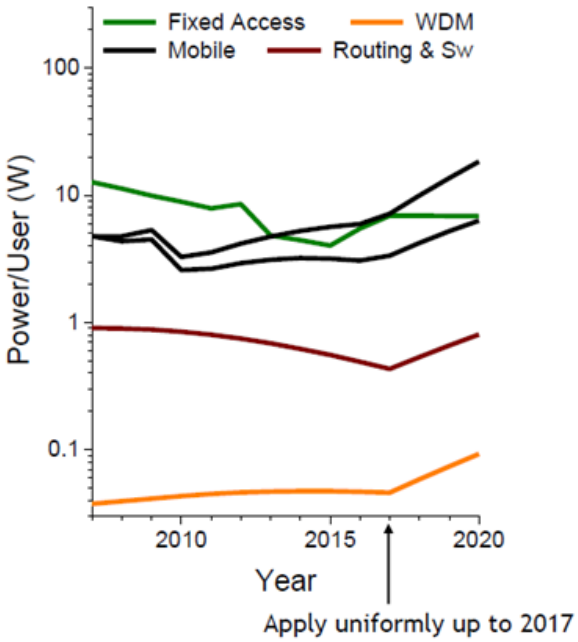

Figure 2.9: Base-Case Efficiency Trends - Energy forecasts assuming optimistic means for efficiency improvements [40].

Network Domain	Component	Capacity	Energy Consumption
Core Network	Core Router (Cisco CRS-1 Multi-shelf System)	92 Tbps	1020 kW
	Optoelectronic Switch (Alcatel-Lucent 1675 Lambda Unite MultiService Switch)	1.2 Tbps	2.5 kW
	Optical Cross-Connect (MRV Optical Cross-Connect)	N/A	228 W
	WDM transponder (Alcatel-Lucent WaveStar OLS WDM Transponder)	40 Gbps	73 W
	EDFA (Cisco ONS 15501 EDFA)	N/A	8 W
Metro Network	Edge Router (Cisco 12816 Edge Router)	160 Gbps	4.21 kW
	SONET ADM (Ciena CN 3600 Intelligent Optical Multiservice Switch)	95 Gbps	1.2 kW
	OADM (Ciena Select OADM)	N/A	450 W
	Network Gateway (Cisco 10008 Router)	8 Gbps	1.1 kW
	Ethernet Switch (Cisco Catalyst 6513 Switch)	720 Gbps	3.21 kW
Access Network	OLT (NEC CM7700S OLT)	1 Gbps	100 W
	ONU (Wave7 ONT-E1000i ONU)	1 Gbps	5 W

Table 2.2: Typical energy consumption values for telecommunication components [79].

demand. A clean-slate approach from multiple angles, including network architectures, protocols (*i.e.* traffic engineering), routing algorithms, transmission systems, and devices (both optical and electronic), is necessary to achieve significant energy savings for a "greener" Internet [80].

Aligned to this energy-focused vision, this work makes some initial strides of using the cross-layer based networking approach to provide a means of optimizing the network's energy efficiency with respect to resource allocation. Using the real-time knowledge of the optical-layer performance and signal degradation provided by

the bidirectional signaling platform, algorithms, network architectures, and optical switching and device technologies are developed that can minimize power while maximizing delivered bandwidth. The energy-efficient allocation of optical network resources is realized with the goal of maintaining application-specific QoS constraints. The goal of this line of work is to allow future routers and switches to be aware of physical-layer impairments in a cross-layer way to reduce the total energy consumption.

This involves close collaborations with GreenTouch, whose mission is to develop an architecture to increase the network energy efficiency by a factor of 1000 from current levels by 2020 [74]. Within the scope of GreenTouch, this work – both current and future – proposes to investigate cross-layer protocols, algorithms, architectures, *etc.* that will increase energy efficiency while leveraging real-time knowledge of the optical layer, specifically geared towards fast IP-layer restoration. IP restoration may be achieved using optical IP switching to correlate optical switching and IP routing within a holistic architecture [81], or through multi-layer protection techniques that can allow for the cost-effective allocation of the physical layer while minimizing the total protection costs [82].

These energy-aware cross-layer approaches will enable more efficient means of resource allocation of the optical network using baseline energy metrics. This may include using sleep modes [79] to conserve router power consumption, since one may note from 2.2 that routers consume a tremendous amount of energy (on the order of kilowatts). The option of optical bypasses [77] may also be considered, whereby the costly O/E/O conversions at each router node may be avoided. Advanced energy-aware traffic engineering schemes [83, 84] will also be considered. Further, OPS may help

play a key role in achieving huge energy efficiency gains, especially with the adoption of photonic integration [85]. The central endeavor of the author's work – which will extend beyond the writing of this dissertation – is to allow the energy-savings techniques to be executed such that there is no disruption of service guarantees, *i.e.* allowing the QoS classes of the network clients to be taken into consideration in conjunction with energy metrics and QoT measurements.

Chapter 3

Optical Switching Fabric Architecture and Test-Bed

This chapter provides an overview of the optical packet switching architecture used in this work, discussing the architecture and an experimental implementation. The optical switching fabric constitutes a major component of the cross-layer box. A technique for increasing the components' switching speeds is given, as well as a means of optimizing the gain uniformity of the switching devices.

To meet the exponentially growing network traffic, future high-capacity data-centric networking systems will need to engage advanced optical technologies. The cross-layer platform investigated in this work will require fast hybrid optical/electronic switches that can be seamlessly and transparently integrated with real-time performance monitoring modules. Researchers have proposed hybrid optical solutions [86], as well as means of packaging the required high-speed integrated devices [87]. This work specifically leverages a fast hybrid optical/electronic switch that relies on a programmable all-optical switching fabric.

The proposed physical-layer switching fabric technology here is OPS, which is advantageous in creating a low-power, low-latency photonic transport infrastructure that can establish end-to-end high-bandwidth lightpaths in a scalable fashion. OPS is a potential solution for realizing routers' high-performing switching fabrics, achieving high bandwidths through WDM to support line-card rates at extremely high data rates. Figure 3.1 shows how the OPS node will fit into the overarching cross-layer platform vision. Future routers that can leverage optical packet switches can achieve a truly adaptable optical layer which is highly reconfigurable and cost effective. Interesting and relevant work [48] has shown that OPS routers can flexibly achieve petabits-per-second throughput while still providing adequate QoS performance.

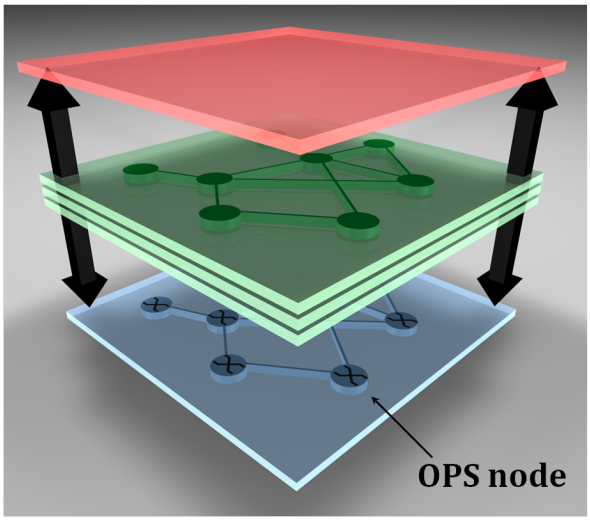

Figure 3.1: Network Architecture with OPS Node - Block diagram of cross-layer protocol stack with integrated OPS nodes in the physical layer.

3.1 Architecture

The OPS architecture designed here has been demonstrated to optically interconnect access/aggregation network edge users or computing network ports; this has been discussed by Shacham *et al.* [25, 88], and an experimental implementation has shown multi-terabit capacity while supporting wavelength-striped, multiwavelength optical packets [26]. The fundamental architecture is based on a transparent multistage fabric topology; the building blocks for the fabric are 2×2 wideband non-blocking bufferless photonic switching elements (PSEs) (Figure 3.2) [89]. From hereafter, the 2×2 PSEs will interchangeably be referred to as photonic switching nodes and photonic switching elements. The architecture uses electronic and optical components with the vision that the complete fabric may be fabricated and integrated on a single photonic integrated circuit (PIC).

Each of the 2×2 PSEs transparently switches the supported optical messages using four fast semiconductor optical amplifier (SOA) gates. The SOA switching gates are organized in a gate-matrix structure. The SOAs support a wide frequency band for transmission, roughly spanning the International Telecommunication Union (ITU) C-band (approximately 1530 nm to 1565 nm). The SOAs additionally provide transparency to the optical packets' data format and bit rate, packet-rate granular switching, sub-nanosecond switching speeds, high extinction ratios, and inherent optical gain to compensate for passive losses. SOAs are a viable approach to meeting the rapid switching requirements of OPS systems (typically in the nanosecond time scale). In the majority of the experimental demonstrations, the supported optical messages are on the order of hundreds of nanoseconds, which span several meters.

Since the messages are longer than the switching elements, no storage or buffering is available within the PSEs (*i.e.* no fiber delay lines (FDLs) are used).

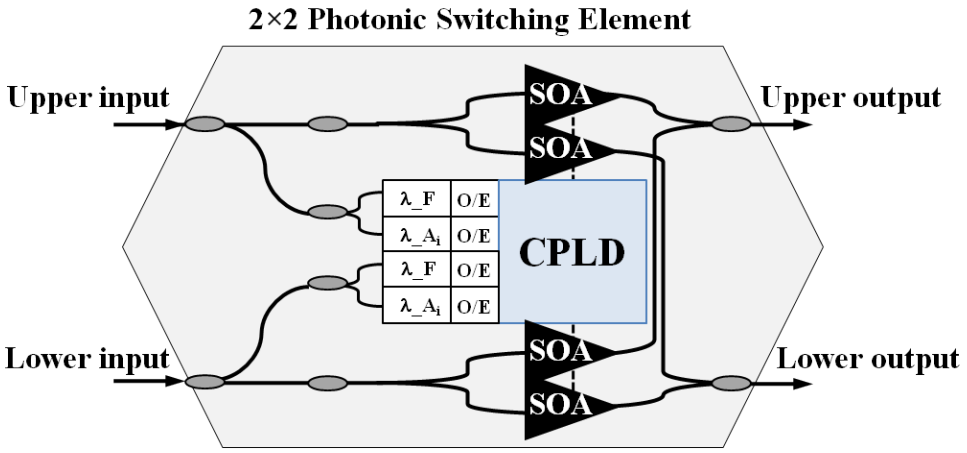

Figure 3.2: 2×2 Photonic Switching Element - Schematic of the 2×2 photonic switching element building block.

Multiple PSEs are interconnected to create a multistage fabric topology. Figure 3.3 provides a typical straightforward example of how the PSE building block structures may be arranged to realize a two-stage, 4×4 switching fabric to connect four independent input and output ports of a router (*i.e.* four line cards). The electronic control logic, synthesized within the PSEs' complex programmable logic devices (CPLDs), provides a high level of programmability to reconfigure the physical connections between PSEs. In future implementations, FPGAs may also be used to provide the routing logic. The basic topology leverages a multistage binary banyan design (specifically an Omega network), that requires $log_2(N)$ of identical stages to create a $N \times N$ interconnect to map a large number of ports [90]. Each stage consists of $N/2$ photonic switching elements, connected in a perfect-shuffle arrangement. In the

simple topology in Figure 3.3, the 4×4 switching fabric requires $log_2(N) = 2$ stages of $N/2 = 2$ PSEs (*i.e.* $N = 4$). The basic PSE has six allowed switching states: (1) the bar state; (2) the cross state; (3) upper straight; (4) upper interchange; (5) lower straight; and (6) lower interchange (Figure 3.4).

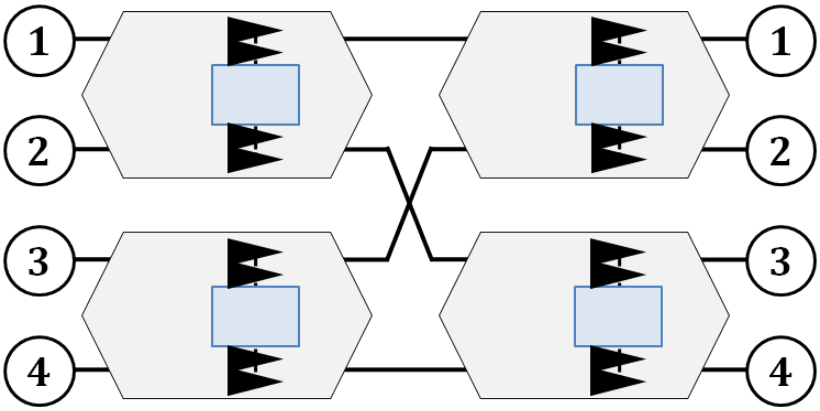

Figure 3.3: Possible Switching Fabric Topology - Example of how the PSEs may be connected in the case of a two-stage, 4×4 fabric implementation.

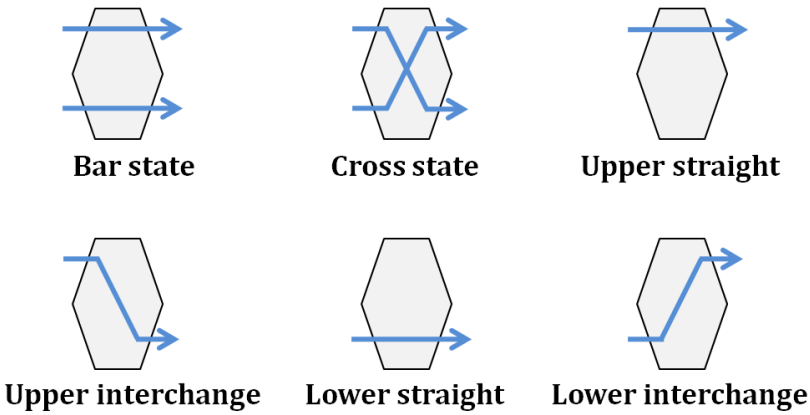

Figure 3.4: PSE Switching States - Configuration of the six switching states supported by the PSE.

The implemented architecture has been shown to support both synchronous [26] and

asynchronous packet transmission [91]. This flexibility alleviates the need to provide complex synchronization tools or modules for the switching fabric; further, since the switching fabric is positioned within a local network node, the issue of clock distribution is mitigated in the case of asynchronous routing. Asynchronous operation is discussed in Chapter 4. Under the assumption of synchronous operation, the switching fabric supports predetermined timeslots with fixed-length packets. At the start of each timeslot, each fabric terminal can begin transmission without prior acknowledgments or requests from a centralized controller. Messages are injected using the input terminals to the fabric, ingressing via the independent input ports, and are transparently and all-optically routed at each PSE.

The packets leverage the abundant bandwidth offered by WDM to provide high transmission bandwidths within the message structure and payload. Using the wavelength-striped packet format, all the wavelengths are routed together as a cohesive whole from the input to the output of the fabric (Figure 3.5). Here, the architecture supports wavelength-striped optical packets, wherein the control header information (*i.e.* the frame, address, and QoS bits) is encoded on a subset of dedicated frequencies, modulated at a single bit per wavelength per timeslot. The packet's control header includes a frame signal F, denoting the presence of a packet and spanning the entire length of the packet; address signals (represented by A_i, A_j in Figure 3.5) denoting the packet's destination used for routing; and a QoS information bit (if required), denoting the packet's priority class (as indicated by a higher-layer protocol). By allowing the control wavelengths to remain high for the duration of the optical message, the switching state of the PSE remains constant as the message propagates through

the switching node. Simultaneously, the packet's payload data is then fragmented and modulated a high data rate (*e.g.* 10 Gb/s, 40 Gb/s, or potentially higher – per data payload channel) on the rest of the supported wavelength band; various modulation formats can be supported. This wavelength-striping approach allows the message to achieve high aggregate transmission bandwidths by allocating the message data to parallel wavelengths that simultaneously contain payload data. The OPS design enables a fast header processing that allows the message to capitalize on the abundant available frequency spectrum provided by the wideband SOAs. Each 2×2 photonic switching element uses a single header bit, which can be a scalable way to achieve a switching fabric with many ports [92]. The message header is such that it may be instantaneously decoded at each PSE and the routing control decision can be made upon reception of the packet's leading edge. The PSEs' electronic control logic is distributed among the PSEs using high-speed programmable circuitry logic, yielding a high level of routing flexibility.

The entire message, including the header and payload, is concurrently routed through the PSEs. At each of the 2×2 PSEs, the actual routing decision is based on the control header extracted from the packet. The leading edge of the optical packet is detected and received at one of the input ports. The framing and address bit signals are extracted immediately using fixed wavelength filters and low-speed *p-i-n* optical receivers. The PSE's routing decision is based on the information encoded in the optical header, which is recovered from the incoming optical packet and processed by high-speed electronic circuitry. The CPLD electronically drives the appropriate SOA gates, and the optical messages are then routed to their desired/encoded destination

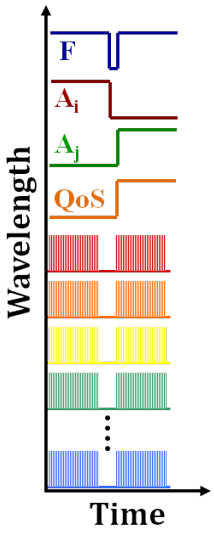

Figure 3.5: Wavelength-Striped Packet Format - Depiction of the wavelength-striped optical packet structure.

(or dropped upon contention). Thus, the PSE's routing decision is based solely on the information encoded in the packets' header. The message payload data and routing control headers are transmitted concurrently to the PSEs and propagate together end-to-end in the switching fabric. The routing logic is distributed among the PSEs and no additional signals are exchanged between them. Each switching element uses simple, combinational logic, and a central fabric control and management plane is not necessary to enable basic photonic switching. No additional signaling is required between the PSEs in a fabric, nor do the elements add (or subtract) information to/from the optical messages. The payload information is not decoded by the PSE logic and is simply routed transparently using one of the four SOA gates. The use of reprogrammable logic devices (*i.e.* the CPLDs) results in straightforward reconfigurability and in the potential for supporting different routing protocols and logic.

No optical buffering is realized within the PSEs; hence, packets are dropped in the case of message contention within the fabric. Though this fabric design is blocking, the topology is significantly advantageous since the individual switching elements can be simply realized at low cost, without the added complexity of optical buffers or wavelength converters. Since each PSE (and consequently each routing stage) has identical propagation delays, the leading edges of messages injected in the same timeslot reach the PSEs simultaneously. Successfully routed messages set up end-to-end transparent lightpaths between fabric terminals. A consecutive series of packets propagating through the test-bed comprise an optical data flow. If the experiment requires this functionality, a physical-layer control acknowledgement (ack) mechanism is implemented whereby short optical pulses are sent to the transmitting node to indicate successful transmission. Due to the instantaneous signaling nature of this protocol, sources that do not receive acks can retransmit synchronously at the next timeslot, yielding a low latency penalty associated with retransmission. The ack scheme is an optional capability of the switching fabric and is not implemented in every experimental demonstration.

3.2 Experimental Implementation

The test-bed environment established by the author features an experimentally implemented optical switching fabric that is based on the aforementioned OPS architecture. The complete test-bed contains twelve realized 2×2 PSEs that can be connected and reconfigured according to the needed experimental functionalities. The basic architecture which is used by the majority of experiments is a 4×4 optical packet

switching fabric, using four 2×2 PSE building blocks (or six PSEs in the case of an enhanced Omega topology).

Each PSE is implemented with discrete macro-scale commercially-available off-the-shelf components, including packaged SOA devices, passive optical devices and couplers, fixed wavelength filters, low-speed 155-Mb/s *p-i-n* photodetectors, and the required electronic circuitry. Figure 3.6 provides a photograph of the realized hardware associated with one PSE node. The high-speed electronic decision logic is synthesized in CPLDs from Xilinx. Each 2×2 switching element uses four SOA gates, designed as a broadcast-and-select topology and organized in a 2×2 matrix. The PSE is capable of decoding optical control bits and maintaining a routing state based on the extracted headers while simultaneously handling wavelength-striped data transparently in the optical domain. Each PSE decodes four control header bits (two for each input port); at each switching stage, the wavelength-based routing information is extracted. The CPLD uses the header bits as inputs in a programmed routing truth table, then gates on the appropriate SOAs. At each 2×2 PSE, the extracted frame bit denotes the presence of a wavelength-striped packet; then, according to the detected address signal, the CPLD will gate the suitable SOA for the packet to be routed to the upper (or lower) output port of the 2×2 PSE (Figure 3.2).

In a typical experimental setup, a 4×4 optical packet switching fabric uses two or three stages, with each stage requiring two PSEs. Figure 3.7 provides a photograph of the hardware for a three-stage fabric using six independent PSEs, realizing an enhanced Omega network. For a 4×4 topology, a minimum of two stages is required for completely interconnecting all output ports with a single possible route. A third

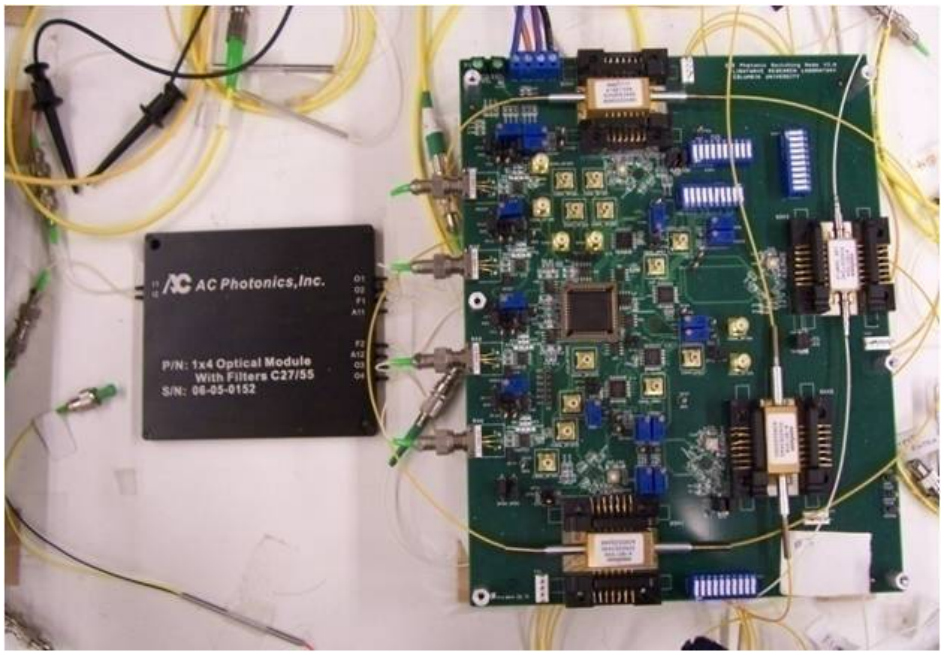

Figure 3.6: PSE Photograph - Photograph of one implemented PSE, showing the passive optical components, electronic circuitry, and SOAs.

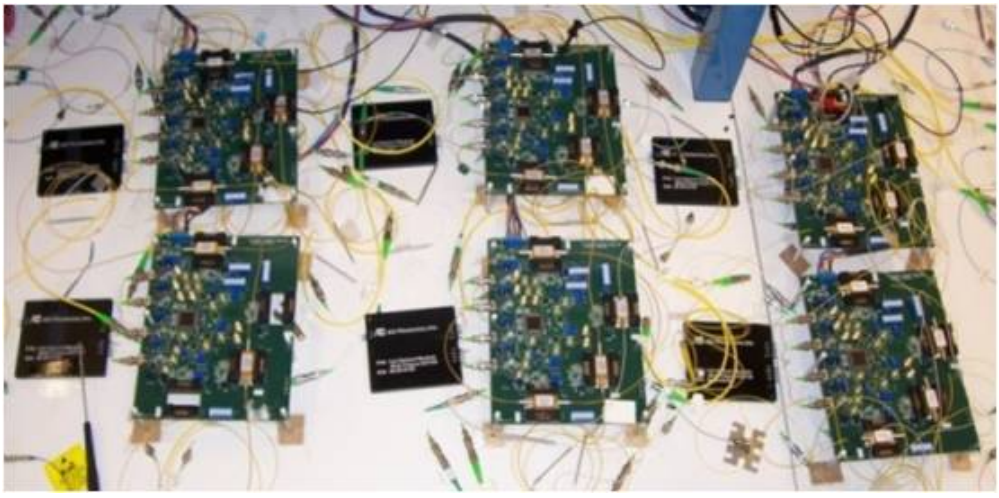

Figure 3.7: Photograph of Optical Switching Fabric - Photograph showing the hardware for a representative three-stage switching fabric implementation.

stage can be used to act as a distribution stage, providing increased path diversity in the case of message blocking and contention within the fabric. The distribution stage allows for two possible routes between any input and output port. The PSEs that share a single fabric stage also share the same address bit. For the current 2×2 PSE, each node filters a two-bit control header for routing: one frame and one address bit. The combinational logic synthesized in the CPLD uses the two-bit control header as follows: upon the presence of the frame bit (F), the logic then examines the address bit. If the address bit is low, the message is directed to the upper output port; if the address is high, the message is transmitted to the lower output port. The PSE hardware supports the simultaneous detection of messages ingressing on two input ports. In these experimental demonstrations, the frame bit is located at 1555.75 nm (C27 within the ITU grid); the possible realizable address wavelengths are: 1531.12 nm (C58), 1533.47 nm (C55), 1535.04 nm (C53), 1543.73 nm (C42), 1550.92 nm (C33), and 1552.52 nm (C31).

The SOAs are operated in their linear, small-gain regime, and are electrically driven with low currents (approximately 50 mA). The SOAs allow for optical amplification to compensate for the insertion loss of the passive optical devices in the PSEs. No net optical power gain or loss is incurred by the optical message as it propagates through each stage. The SOAs provide a low-power switching gate over a wide frequency band such that thermal variations do not negatively affect performance significantly. During timeslots when the PSEs are not switching optical messages, the electronic control logic does not gate on the SOAs; thus, the PSEs consume negligible power. The majority of the SOAs used in this test-bed are available commercially from Kamelian/Amphotonix

(C-band Optical Power Boosters), though devices from other vendors are also used, and have rated noise figures of approximately 6.5 to 7 dB. Each SOA provides approximately 8.5 dB of optical amplification. The optical powers of the input packets are maintained such that the SOAs do not add nonlinearities to the propagating signal (*i.e.* the average power of the control headers is approximately -8 dBm and the payload channels are each around -16 dBm). The SOAs are mounted on a custom-designed electronic circuit board (Figure 3.6) with the required integrated electronic components and low-speed optical receivers.

3.3 Experimental Packet Generation and Analysis Setup

The exact experimental setups associated with the optical packet generation before the switching fabric and the packet analysis at the output of the fabric differ slightly for each demonstration. In this chapter, a general setup is described and the specific details are given with each of the demonstrations in the following two chapters.

In each experimental demonstration, a pattern of multiwavelength optical messages are generated that exemplify the specific routing functionality that is desired. In all cases of optical packet generation, the payload channels for the wavelength-striped packets are generated using discrete continuous-wave (CW) distributed feedback (DFB) lasers each connected to a polarization controller (PC). The laser signals are combined onto a single fiber using a passive optical multiplexer. All the payload wavelength channels are then simultaneously modulated with a high-speed radio frequency (RF) signal (*e.g.* a 10-Gb/s nonreturn-to-zero (NRZ) signal with an on-

off-keyed (OOK) format) that typically carries a pseudo-random bit sequence (PRBS). A single $LiNbO_3$ amplitude modulator is habitually used to modulate all the payload channels concurrently. The modulator is driven by a high-speed electrical signal from a pulse pattern generator (PPG). The wavelength channels are then decorrelated by a span of single-mode optical fiber (SMF), typically in the order of a few to tens of kilometers long. The payload wavelengths are then split using a passive optical coupler to create several modulated wavelength-striped data flows for injection in the switching fabric ports. Each set of payload wavelength signals is then transmitted to external gating SOAs. The gating SOAs provide additional amplification and help form discrete optical packets. The payload channels are chosen to span the ITU C-band, showing the wideband capabilities of the fabric, and with a minimum spacing between two payload channels of 100 GHz (equivalent to 0.8 nm), demonstrating the lack of crosstalk contributed by the SOAs.

The control wavelengths are generated using separate CW DFB lasers, including one frame at 1555.75 nm and several address bits, ranging from 1531.12 nm to 1552.52 nm. The DFB lasers are split using passive couplers (to create one control bit per injection port) and sent to a set of gating SOAs. The control header and payload data signals are then gated into packets using an array of packet modulation SOAs, encoding the appropriate addressing information for each packet to be routed through the implemented fabric. The control headers and the payload signals are then passively combined together to create a multiwavelength packet stream. A similar packet-generation setup is used concurrently for each set of control and payload signals to form a distinct packet pattern for the each of the input ports of the fabric.

All the gating SOAs for packet gating and fabric addressing are controlled by a fast, nanosecond-scale programmable electronic pattern generator (PG), typically an Agilent ParBERT or Tektronix Data Timing Generator (DTG). The PG is pre-programmed with test packet patterns that are custom-designed for each experimental demonstration. The complete system thus creates wavelength-striped packets with a multi-bit control header and a multiwavelength payload (with each payload carrying a 10-Gb/s or 40-Gb/s bit rate). These packets are then injected into the active ports of the switching fabric. If implemented in a particular experiment, the ack patterns are created such that an ack pulse is transmitted at the same time as when a packet is expected to arrive at the respective output port.

At the output of the realized switching fabric, the multiwavelength packet is monitored and examined using an optical spectrum analyzer (OSA) and high-speed sampling oscilloscope (*i.e.* a communications signal analyzer (CSA) or digital communications analyzer (DCA)). A packet analysis system is also used in which the wavelength-striped packet propagates to a tunable grating filter. The filter selects one payload channel for signal integrity analysis and rejects the accumulated amplified spontaneous emission (ASE) from the SOAs. The payload channel is then sent to an erbium-doped fiber amplifier (EDFA), another tunable filter to reduce the ASE from the EDFA, and a variable optical attenuator (VOA). The payload wavelength channel is then received by a DC-coupled 10-Gb/s *p-i-n* photodetector followed by a transimpedance amplifier (TIA) and limiting amplifier (LA). The received electrical signals are sent to a bit-error-rate tester (BERT) that is synchronized with the PPG (typically, no clock recovery is performed) and gated to analyze the packets with

the ParBERT. The experimental demonstrations show correct functionality of the switching fabric, with correct addressing and switching. BER measurements further validate the signal integrity of the test-bed setup, nominally demonstrating error-free operation (as defined by attaining BERs less than 10^{-12}) and providing a means for power penalty analysis.

3.4 SOA Switching Speed Improvements

This section discusses a possible technique to improve the switching speeds of SOAs using a multipulse current injection technique to support the fast routing of wavelength-striped optical packets [93]. A reduced SOA switching speed is demonstrated for 8×10-Gb/s wavelength-striped optical packets using multipulse pre-emphasis current injection. The scheme yields a 20% improvement in switching speed; the resulting power penalty performance is also characterized.

3.4.1 Multipulse Current Injection for SOAs

SOAs comprise a key optical device for implementing OPS networks, due to its fast switching speeds, high extinction ratios, data transparency, broad gain spectrum, and ability to be integrated on a chip [94, 95]. SOAs may be deployed as amplifiers, add-drop devices in optical links, and wavelength converters in WDM networks. SOAs have also been utilized in optical test-beds (such as OSMOSIS, developed by Corning and discussed in [96], and the data vortex interconnection network, developed by Columbia and detailed in [97]), specifically acting as a fast switching gate within a 2×2 nonblocking wideband photonic switching element in order to achieve programmable

high-speed switching on the optical physical layer [26]. It is thus evident that further reduction of SOAs' switching speeds would be desirable to improve the overall performance of future optical implementations.

It has been previously proposed to realize predistortion and preshaping techniques to enhance the speed of laser diode switches [98]; similarly, pre-emphasis schemes have been used recently to increase the switching speed of silicon ring modulators [99]. A corresponding technique may be adopted to reduce the turn-on response time for SOAs, whereby an additional pre-emphasis impulse current can be injected in the device's active region to improve its rise time performance [100, 101]. Controlling the device's carrier population/depletion properties, and hence the carrier lifetimes, provides improved switching times. Previously, the effectiveness of this multipulse pre-emphasis current injection technique was shown in simulation and in experiment to reduce SOA switching times [101]. The work here shows an improvement in the fast switching of high-bandwidth optical packets using a SOA device with multiple injection currents by transmitting wavelength-striped 8×10-Gb/s optical packets through the SOA. Furthermore, the BER performance of the SOA with and without the multipulse pre-emphasis drive currents is characterized and a proven 0.05-dB improvement in the device's power penalty performance is shown. This work showcases the feasibility of using a multipulse current injection technique for a SOA to reduce the switching time of high-data-rate optical packets, vastly improving the device's deployment as switching elements in future OPS networks.

The SOA switching time improvement scheme is based on a pattern-modification of the injected current (Figure 3.8). The multiple pre-emphasis current injection pulses

consist of a step signal (V_{step}) and an impulse signal (V_1) set to have simultaneous, overlapping rising edges. The scheme is used to incur an increased optical gain at the onset (*i.e.* rising edge) of the step pulse. The implemented shape of the current injection pulses are fast step signals, applied to the SOA in order to turn the device on or off. The on/off response is not instantaneous, as the device's active region must be populated by the injected carriers, or depleted by the suppressed carriers. An additional pulse is used to cause a rapid increase in the carrier population and thus reduce the overall carrier lifetime, and thus by extension, the SOA's switching times. The SOA must remain in its linear regime of operation during the injection of all additional drive currents.

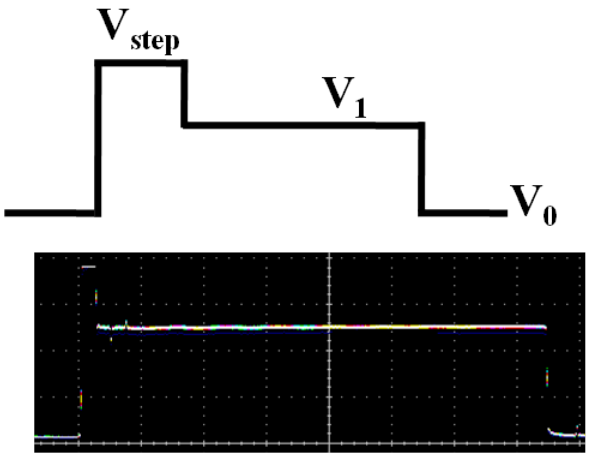

Figure 3.8: Multipulse Current Injection Technique - The implemented multipulse current injection pulse for reducing the SOA turn-on response time.

3.4.2 Experimental Setup

A two-part experiment verifies that this pre-emphasis current injection yields a reduction in the SOA switching time. In the first experiment, one CW-DFB laser transmits a single wavelength to the SOA; an improved response time is achieved. In the second experiment, high-bandwidth 8×10-Gb/s wavelength-striped optical packets are generated and propagate through the SOA, and the BER performance is assessed while the SOA performs a fast switching of broadband WDM optical packets. Since the first experimental setup is a subset of the second, only the latter setup is discussed below.

The second experimental setup (Figure 3.9) for the subsequent BER measurements uses eight CW lasers (DFBs in Figure 3.9) (ranging from 1537.4 nm to 1559.7 nm), which are combined with a passive multiplexer. The eight wavelength channels are simultaneously modulated with a 10-Gb/s NRZ signal that carries a $2^7 - 1$ PRBS using a single $LiNbO_3$ modulator (mod in Figure 3.9). The modulator is driven by an electrical signal from a 10-Gb/s PPG. The wavelength channels are decorrelated by 25 km of optical fiber and then propagate through the SOA. The ParBERT generates the two synchronized electrical injection currents (*i.e.* the step pulse signal and impulse signal), which are then delivered to the SOA as a single combined multipulse drive current pulse. At the output of the SOA, the multiwavelength signal is monitored by an OSA, while one wavelength channel propagates through a tunable grating filter (λ in Figure 3.9), an EDFA, a second tunable filter, a VOA, and is then received by a 10-Gb/s *p-i-n* photodiode with TIA and LA pair (RX). The packet analysis is performed with a BERT that is synchronized with the PPG and gated for packet analysis by the

ParBERT.

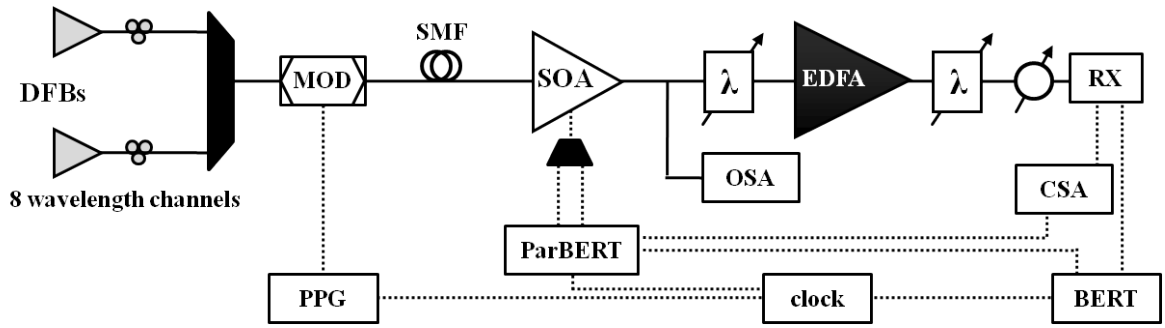

Figure 3.9: Multipulse Current Injection Experimental Setup - Setup used to determine BER performance. Solid lines indicate optical fiber, while dashed lines indicate electrical cable.

3.4.3 Results and Discussion

The SOA used in this experiment is a state-of-the-art Kamelian (Amphotonix) device (specifically, part number: OPB-10-15-N-C-FA) that is set to maintain operation in its linear regime. The impulse signal is set to 5 ns and it is observed that varying the length of the impulse does not alter the SOA's switching speed; however, varying the pulse's amplitude directly affects performance. The step signal is set to 150 ns; for an eight-channel wavelength-striped packet at 10 Gb/s, the aggregate packet size is equivalent to the 1500-byte maximum transmission unit (MTU) of an Ethernet packet.

In the first experiment (*i.e.* using the single CW laser), a reduction in switching time of 0.5 ns is obtained (Figure 3.10a), thereby validating the use of the multipulse current injection technique. In the subsequent experiment, broadband 8×10-Gb/s optical packets propagate through the SOA and its 20/80 rise time performance is examined (Figure 3.10b). The switching on time (*i.e.* rise time) without the pre-

emphasis current injection is approximately 1.1 ns, while the corresponding switching on time with the pre-emphasis current injection is approximately 0.9 ns. Thus, a 20% decrease in device switching time is shown with this technique. The BER performance of SOA with and without the pre-emphasis current injection is also characterized for the 8×10-Gb/s optical packets. Using the ParBERT, the BERT is gated specifically on the starting bits of the packet to quantify the effect of the pre-emphasis drive current. Since the rise time of the SOA is improved, the BER is also improved at the rising edge of the packet. Error-free performance is confirmed at the output of the SOA, achieving BERs less than 10^{-12} on all eight wavelength channels of the wavelength-striped optical packet. Sensitivity curves for one supported wavelength for the device with and without the pre-emphasis signals are shown in Figure 3.11. The SOA's operation without the multiple injected currents exhibits a 0.35-dB power penalty (taken at a BER of 10^{-9}), while the SOA operating with the multiple current injection achieves an improved power penalty of 0.3 dB. Thus, this multipulse current injection technique does in fact yield a better power penalty performance.

One should note that the Kamelian SOA has a rated rise time of 0.9 ns. Here, the SOA switching time is affected by the non-idealities of the experimental implementation. In practice, the SOA packaging and the trace layout extending between the current driver die (here, a commercial laser driver (MAX3656)) and the SOA cathode pin can pose limitations on the response time [102] due to the presence of small capacitances and inductances. Accounting for these practical values, the achieved rise times are reasonable. Further, only the rise time performance of the SOA with the pre-emphasis current injection is studied here; this work can easily be extended

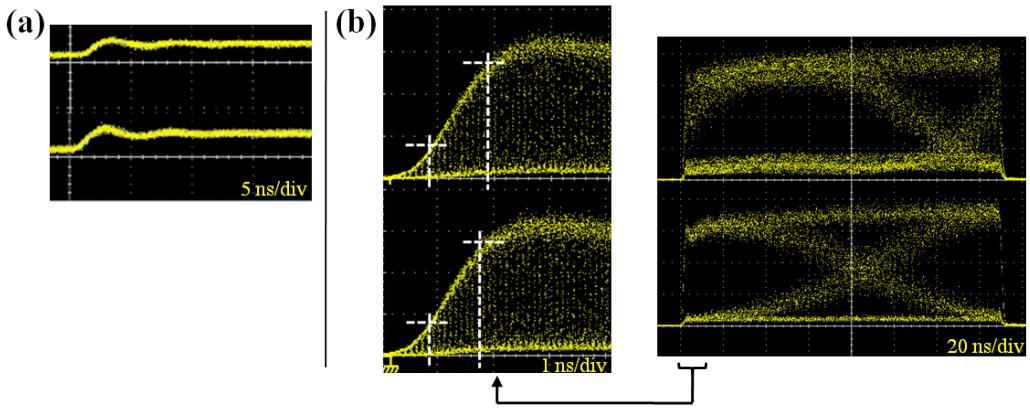

Figure 3.10: Multipulse Current Injection Traces - Waveform traces corresponding to the multipulse injection experiment. The top shows the traces without pre-emphasis current injection, while the bottom depicts results with pre-emphasis current injection. (a) Rise time of a single CW signal; (b) Waveforms of one 10-Gb/s channel of the 150-ns optical packet with magnified views of the rising edge: dashed lines show the 20/80 rise time.

to improve fall (*i.e.* switching off) times. This work provides a relative study to the feasibility of a multipulse current injection for SOAs. In an ideal implementation, a differentiator and an integrated custom-designed pre-emphasis circuit can vastly improve both the SOA switching on and off times.

Thus, a multiple pulse pre-emphasis current injection technique is demonstrated for a SOA device that shows a proven reduction in switching speed. This implementation is validated using the fast switching of 8×10-Gb/s wavelength-striped optical packets. The rise time is reduced by 20% and a 0.05-dB power penalty improvement is shown. This study in optimizing SOA switching may be valuable in the development of OPS in future large-scale optical networks.

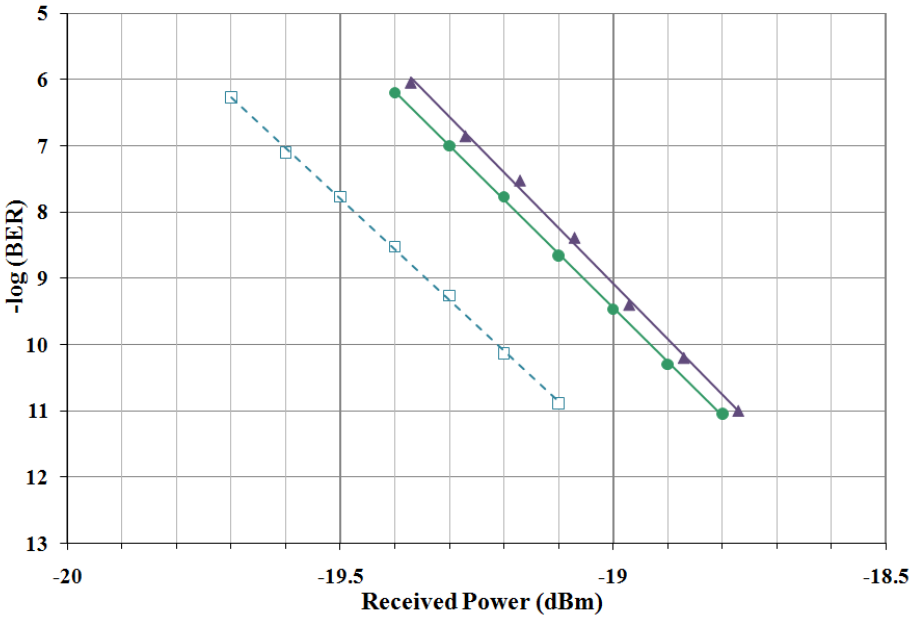

Figure 3.11: Multipulse Current Injection Sensitivity Curves - BER sensitivity plots for one wavelength channel ($\lambda = 1559.7$ nm) recorded without the multipulse pre-emphasis signal [purple filled points], with the multipulse signal [green filled points], and the back-to-back signal [blue unfilled points].

3.5 SOA Gain Uniformity Optimization

With the help of the author, the bandwidth and scalability performance of multiwavelength OPS networks has also been improved using a design methodology for optimizing SOA gain uniformity in Lee *et al.* [103]. This work addresses the design of an OPS switching node (*i.e.* the PSE) when utilizing SOAs as broadcast-and-select switching components. Experimentally, it is shown that the SOAs can achieve spectral uniformity when operated near a specific drive current value, which is distinct to the SOA device structure. The methodology is explored and validated using a commercial Alcatel device operating under the proposed conditions, showcasing its performance with respect to gain, noise figure, OSNR, and power penalty.

As discussed in [103], for OPS networks that leverage SOAs as wideband switching components, the robustness of the switching node (the PSE) is limited by the curvature of the SOA's gain spectrum. In order to achieve a more uniform performance of the different wavelengths within a WDM packet, it is proposed that the SOA can be operated using an optimal drive current that delivers the most uniform gain. Experimentally, five signal wavelengths that span the C-band are transmitted to the SOA. It is found that as the drive current increases, the wavelength that experiences that highest gain decreases. By sweeping the SOA drive current and recording the gain of each channel, the most uniform-gain operating point can be determined experimentally for each SOA device. Similarly, the noise figure of the SOA can be characterized with respect to the drive current.

The power penalty and OSNR performance of the device can also be investigated when operating the SOA under this predetermined operating point. The power penalty

metric is used in establishing a system-level optical power budget, since it is defined as the amount of power needed at the receiver to overcome the bit errors introduced by the system or device under test [16]. SOA operation at the drive current that provides the optimal gain uniformity allows high OSNR values to be maintained after undergoing four SOA hops. The gain variance is reduced at no additional cost of the worse-case power penalty; this is meaningful to the network's scalability performance.

Thus, this section of the thesis summarizes the work in [103], which proposes a design scheme for optimizing the performance of SOA-based OPS elements. Using a detailed *a priori* characterization of each SOA in terms of gain and noise figure, the performance of OPS networks can be improved by determining the optimal operating point to support multiwavelength optical packets. Optimizing the gain uniformity of these SOA components can yield improvements in network performance and scalability.

3.6 Discussion

As a final note, it must not be neglected to mention some of the arguments against the development of OPS. Today's electronic routers assume a store-and-forward scheme with electronic random access memories (RAMs). To enable the widespread deployment of optical routers, the lack of a practical optical RAM technology is recognized as a significant challenge that must be overcome [23]. The overarching flaw with this approach is that it assumes a one-to-one replacement of electronic router components with equivalent optical counterparts. This paradigm will not yield optical routers that can truly benefit from the advantages of photonics. Indeed, hybrid approaches will likely be required (*i.e.* incorporating both electronic and

optical devices to exploit each domain's advantages). The author proposes here a hybrid switching fabric solution: the routing is performed electronically, leveraging the programmability of the electrical domain, while the data is carried optically, utilizing the photonic domain for its high-bandwidth capabilities. Furthermore, others maintain that electronic buffering will remain the technology of choice in future high-capacity routers [104], and that adopting hybrid OPS systems (using electronic RAM, thus without optical buffers) can allow for improvements in energy efficiency [85]. Hybrid switches have also been shown to be better performing than fully electrical and all-optical counterparts [105].

Further, it should be noted that the SOA devices used here provide a convenient platform for the switching functionality and experimental testing. The author acknowledges that SOA components themselves do not exhibit low-power performance, and thus will likely not be the physical-layer technology of choice for future energy-efficient optical switching.

Chapter 4

Advanced Optical Switching Functionalities

This chapter outlines the high-level switching functionalities that have been developed for OPS fabrics and have been demonstrated by the author in the previously-discussed optical networking test-bed.

The ultimate goal of this body of work is to achieve greater functionality in the optical switching fabric. Service providers, network operators, and researchers all agree that by driving more functionality from the electrical layer to the lower optical layers, greater cost savings may be achieved [39]. By allowing functionalities that are currently executed electronically by higher network layers to be realized at the optical transport layer, both operational (OpEx) and capital expenditures (CapEx) can be reduced. Since transport is inherently less costly lower in the OSI network stack, this work aims to provide the physical layer with more dynamic functionalities and capabilities. Greater automation at the optical layer is required in order to reduce the cost of provisioning bandwidth. By migrating more of the network's functionality to the

optical layer, traffic can be switched more effectively at the physical/optical layer. This perspective motivates all-optical switching for future networks with minimal O/E/O conversions, and the effects of this thinking can clearly be seen in recent innovations in next-generation reconfigurable optical add-drop multiplexers (ROADMs).

Thus, the following chapter of this dissertation addresses achieving more high-level, advanced network functionalities at the optical switching fabric layer. This will bring a unique and intrinsic dynamicism to the physical layer by supporting highly reconfigurable switching techniques, as well as an advanced management and handling of optical packets entirely in the photonic domain. The following capabilities have been developed by the author and demonstrated using the implemented optical switching fabric test-bed: the support for asynchronous routing, optical QoS based routing, the capacity for multi-terabit transmission, in addition to several explorations on wavelength-striped packet multicasting for optical switching fabrics. These functionalities are important features for the cross-layer enabled network node.

These advanced optical switching capabilities can be straightforwardly extended beyond the scope of future routers and also be applicable to optical interconnects within high-performance computing (HPC) systems. Current interconnection networks are being increasingly constrained by the limited bandwidth, latency, and power performance of electronics. Optical interconnection networks (OINs) have been suggested as a promising solution to these performance bottlenecks [96, 106]. Greater functionality and networking capabilities of the OIN comprises an important development to further the feasibility of optical interconnects in HPC systems.

4.1 Asynchronous Operation

The asynchronous operation of routers (and potentially of deployed OINs in future HPC infrastructures) is a key functionality to enable the scaling to larger network sizes. The switching fabrics within the router may be required to operate asynchronously without the need for a synchronization stage or timeslot alignment module. This may be especially important in the case in which multiple independent network nodes are deployed, each featuring its own clock.

Asynchronous operation provides several advantages as compared to synchronous transmission, with no required temporal alignment between optical messages entering the fabric from different ports. This architectural approach yields two key benefits: first, synchronizing all the transmission ports with a common clock may be costly. There is an inherent timing jitter between the global clock and the possible packet-level clock, due to slight frequency drifts [107]. Asynchronous operation alleviates the need for fine temporal alignment, calibration, and synchronization. Second, in asynchronous mode operation, variable-length packets are supported, providing greater flexibility in terms of traffic and enabling the exchange of small optical packets. This may be acutely important for certain applications which may use short control packets with little data. In synchronous networks, the exchange of these messages using full timeslots can lead to significant underutilization [108]. Thus, asynchronous transmission can lead to more flexible operation and allow for more flexible scheduling [109].

There have been several approaches to achieving asynchronous packet routing. Some related work has been shown to align incoming packets from multiple sources to the fabric's timeslots using packet synchronizers and optical buffers [110]. The

end-to-end asynchronous optical packet switching and forwarding has also been shown for IP packets supporting 12.5-Gb/s payloads with 3.125-Gb/s controls [111]. This was demonstrated on a wavelength-routed network using a SOA based Mach-Zehnder interferometer (MZI) wavelength converter, tunable sources, and arrayed waveguide grating (AWG).

In this effort, the asynchronous transmission demonstration leverages the simple reprogrammable capabilities of the implemented optical switching fabric and requires no additional hardware to support asynchronous routing [91, 112]. The basic architecture was originally developed as a synchronous slotted network, transmitting fixed-duration optical messages. The architecture can be straightforwardly adapted to provide asynchronous operation by incorporating simple, minimal modifications to the PSEs' electronic circuitry. The fabric yields simple asynchronous packet switching without requiring control plane signaling, optical buffering (*i.e.* FDLs), or wavelength conversion. No additional hardware or components are needed to implement the asynchronous transmission as compared to the synchronous case. Unlike designs that require a centralized arbiter to manage routing, the switching fabric's unique distributed control nature is leveraged. The issue of signaling between the PSEs/nodes and a centrally controlled arbiter, or between the input terminals and central arbiter, is alleviated.

This section presents the experimental demonstration and performance evaluation of asynchronous transmission of arbitrary-length, wavelength-striped optical messages across a three-stage, 4×4 OPS fabric test-bed. This allows for the asynchronous routing of multiwavelength optical packets through an optical network without additional

hardware. Wavelength-striped optical packets with 6×10-Gb/s payloads are correctly routed through the test-bed, and error-free transmission with BERs less than 10^{-12} is confirmed for all payload wavelengths. Sensitivity curves show an average induced power penalty of 0.5 dB for the six payload wavelengths.

4.1.1 Asynchronous Demonstration

The three-stage, 4×4 switching fabric test-bed is implemented as discussed in Chapter 3. Figure 4.1 shows the fabric topology and corresponding test-bed photograph. Here, the fabric within the test-bed is adapted to asynchronously route variable-length messages by simply incorporating a two-bit electronic memory register to encode the switching state in each PSE's electronic routing logic. Under the assumption of synchronous operation, all messages are received simultaneously at the beginning of the timeslot; thus, simple stateless combinatorial logic is sufficient to extract and process the control header and route optical messages (*i.e.* gate on the appropriate SOA). However, when messages arrive at differing times in an asynchronous fashion (and with varying lengths), the two-bit memory is needed to denote the PSE's state in order to give priority to lightpaths that have been previously set and to avoid interference with new messages. Optical packets that arrive first are given priority in fabric transmission. Each PSE is set on a per-packet basis according to the headers of the ingressing messages and the state of a given PSE is independent of other PSEs.

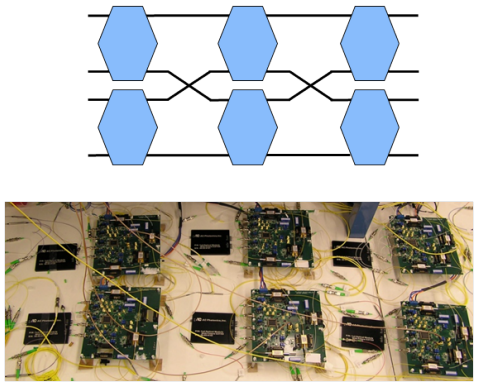

Figure 4.1: 4×4 Switching Fabric Topology - Depiction and photograph of three-stage implemented switching fabric.

4.1.2 Experimental Results

In order to verify the fabric's straightforward capability to route asynchronous traffic with variable packet lengths, a pattern of wavelength-striped packets is injected into the test-bed using three independent input ports. Figure 4.2 depicts the optical waveforms corresponding to the input and output pattern signals.

The optical packets (labeled A-G in Figure 4.2), with lengths varying between 53.3 ns and 409.6 ns, are injected from three input ports (in0, in1, and in2). There is no assumed relationship between the individual packets' start and end times, and no timeslots are necessary. Each packet occupies the full duty length, incorporating a four-wavelength control header and six payload wavelengths. CW-DFBs are used to generate the incoming messages. All payload wavelengths are simultaneously modulated at 10 Gb/s with a $LiNbO_3$ amplitude modulator with a 2^7-1 PRBS in a NRZ-OOK format. The payload wavelength channels in this experiment range from 1539.6 nm to 1558.28 nm. On the dedicated control wavelengths, the messages have

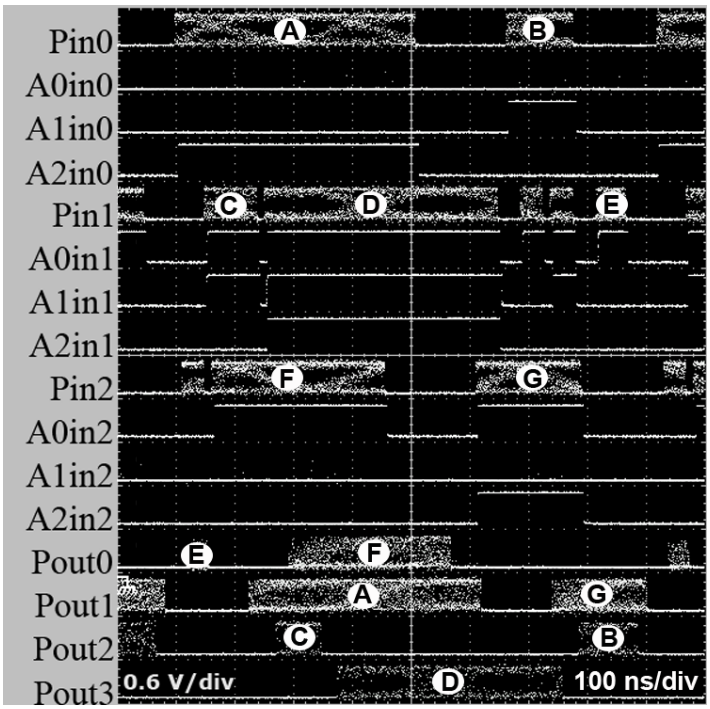

Figure 4.2: Asynchronous Operation Traces - Experimental optical waveform traces associated with the asynchronous traffic.

optically encoded addresses denoting the designated output (out0, out1, out2, and out3), modulated at a single bit per wavelength. The optical header has one frame signal (not shown in Figure 4.2), a distribution address (A0, selecting one of two paths in the three-stage test-bed), and a two-bit routing address (A1, A2). For example, packet C (which is 89.2 ns long) is injected from in1, addressed to out2 (A1=1, A2=0), and emerges at out2.

Ack pulses are implemented in this work. Following injection, a packet may be blocked by other packets whose transmission began earlier. The new packet is dropped and its source does not receive an ack. In the case of no ack, the source recognizes that the packet was blocked and retransmits via a different path. As an example, the first injection attempt from in2 is blocked; the source retransmits with a different distribution address (*i.e.* A0 changes from 0 to 1 in the subsequent transmission) and another path to the destination is found (packet F). This demonstration capitalizes on the increased (2×) path diversity provided by the enhanced Omega design.

At the output, an optical filter selects one payload wavelength channel, which is then sent to a VOA, and subsequently to a direct-coupled 10-Gb/s *p-i-n* receiver with TIA and LA pair. A BERT is used that is synchronized with the packet gating signals. Error-free routing is verified for all six payload wavelengths of the egressing asynchronously-routed packet, achieving BERs less than 10^{-12}. Figure 4.3 shows the 10-Gb/s input and output eye diagrams, with no extinction ratio difference. The power penalty induced by the test-bed is evaluated for all six payload channels. Figure 4.4 shows a representative set of 10-Gb/s sensitivity curves for one of the evaluated wavelengths, validating the fabric's error-free performance. Figure 4.5 gives the power

penalty (at a BER of 10^{-9}) as it relates to wavelength. It is observed that the average power penalty is approximately 0.5 dB for the three-stage fabric, ranging from 0.3 dB to 0.8 dB. The implementation is not limited by the SOA's ASE. It is significant to note that the obtained power penalty values are less than 1 dB for all payload channels. The power penalty values verify that this approach to asynchronous routing does not adversely affect the fabric's operation, as compared to synchronous transmission.

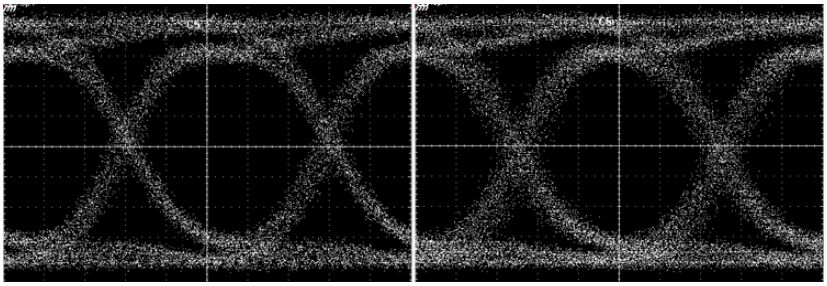

Figure 4.3: Asynchronous Operation Eye Diagrams - 10-Gb/s input (left) and output (right) eye diagrams corresponding to the asynchronous demonstration (λ=1560.2 nm)

Thus, using minimal electronic logic circuitry modifications, the asynchronous mode operation of an implemented optical switching fabric in the test-bed is successfully demonstrated with variable-length wavelength-striped optical packets. Multiwavelength optical packets with 6×10-Gb/s payloads are shown correctly routed asynchronously through the fabric. The correctly routed packets deliver six wavelength channels of 10-Gb/s payloads with confirmed error-free, with BERs less than 10^{-12}. An average induced power penalty of 0.5 dB is shown for the six payload wavelengths. The achieved power penalties for asynchronous transmission are similar to the original slotted operation. This experimental demonstration paves the way for the enhanced, scalable performance of optical switching fabrics in future routers.

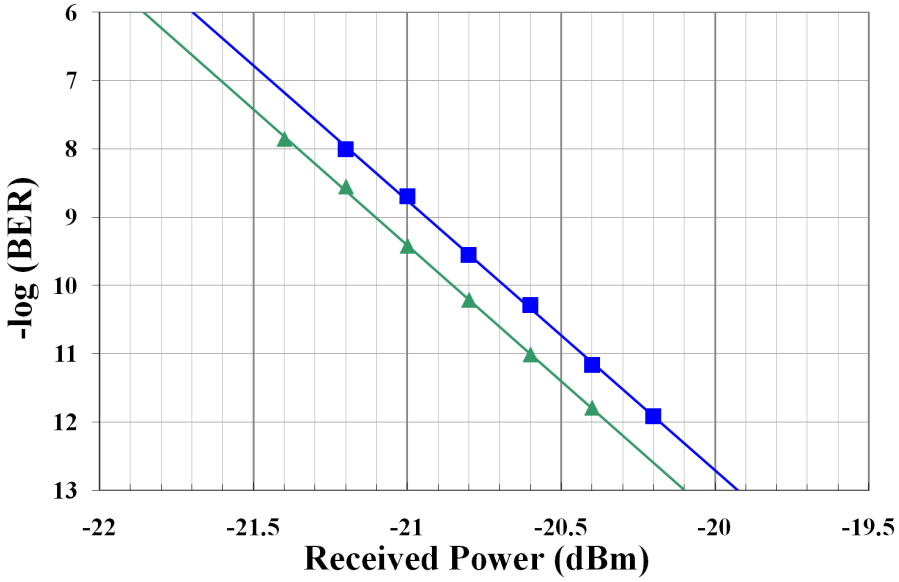

Figure 4.4: Asynchronous Operation Sensitivity Curves - 10-Gb/s BER sensitivity curves for one representative wavelength ($\lambda = 1558.28$ nm).

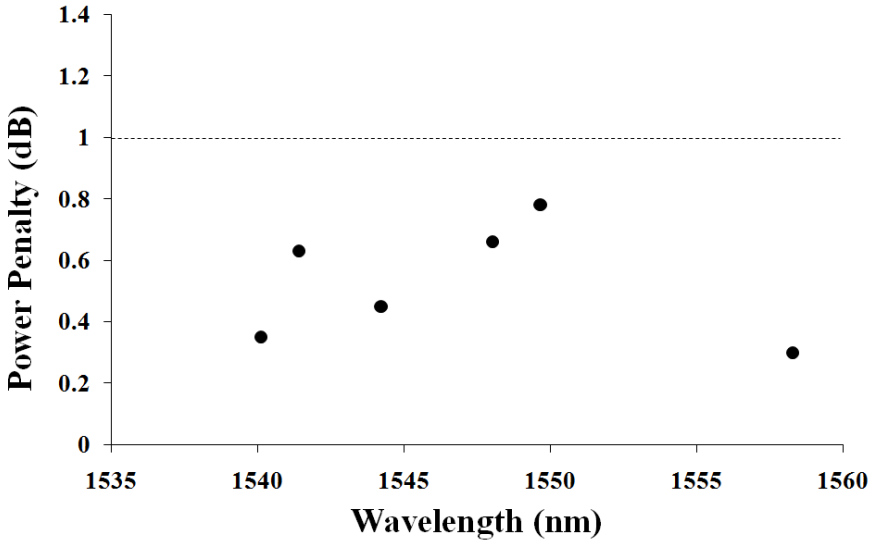

Figure 4.5: Asynchronous Operation Power Penalties - Power penalty values with respect to all six payload wavelength channels.

4.2 Optical Quality-of-Service Based Routing

This section presents a novel encoding scheme that allows application-specific QoS requirements to be mapped directly onto optical packets in the physical layer. The optical QoS (OQoS) scheme is discussed and its implementation in the switching fabric test-bed is described [113, 114].

Cross-layer optimized routing algorithms may be required to invoke QoS classes directly on the optical layer to optimize end-to-end global network routing, as well as exploit OPS architectures as a solution to switch high-bandwidth optical messages. In this way, QoS classes can then be leveraged by higher-layer applications to optimize optical-layer routing. Currently, IP-layer QoS constraints are not adequately supported by the lower layers (including the physical layer); by guaranteeing QoS directly on the optical layer, improved network performance may result with increased support of best-effort and high-priority requirements. These mechanisms for QoS provisioning must also account for the physical-layer impairments, *i.e.* the cross-layer platform should be aware of a packet's OQoS during message transmission and routing [27, 115].

OPS fabrics may offer several higher-quality or higher-priority connection-oriented services [116]. This may help alleviate the significant challenge of contention resolution within the fabric, which may be required as multiple packets attempt to egress simultaneously on the same link. Contention resolution is not easily managed, owing to the currently infeasible realization of optical buffers. Current schemes use packet-dropping and retransmission, which may be costly for important messages. Implementing QoS classes on the optical layer may mitigate the expensive issue of contention resolution in OPS fabrics [117].

Further, the optical network should support user-differentiated protocols embedded on the physical layer through varying levels of QoS and priority. QoS-aware routing schemes for optical networks have been presented for OBS networks [118] and the loss performance of a multi-QoS scheme has been studied in simulation for OPS networks [38]. OPS fabric performance may be significantly improved by realizing service classes through different packet priority levels and by the routing of prioritized optical messages. In order to adequately implement these service-aware routing schemes, an optimal coding mechanism must be experimentally demonstrated to show the feasibility of creating QoS-aware optical messages.

Here, an OQoS encoding scheme is introduced for an optical switching fabric testbed that allows for the prioritized transmission of broadband wavelength-striped optical messages. The switching of optical packets accounts for a diverse set of OQoS classes that are encoded directly within the optical messages. The OQoS priority encoding mechanism specifically addresses contention resolution in future OPS fabrics [25, 114]. Contention is resolved by dropping low-priority messages, which can be retransmitted in a subsequent timeslot. This OQoS-aware routing scheme prevents high-priority packets from being dropped, allowing for an overall reduction in packet retransmission penalty for critical data streams. 8×10-Gb/s wavelength-striped messages are shown correctly routed error-free, with verified BERs less than 10^{-12}. A power penalty of 1.2 dB for a three-stage synchronous optical fabric is demonstrated.

4.2.1 OQoS Encoding Scheme

Due to the straightforward reprogrammable capability of the PSEs' control logic, the implemented OPS fabric test-bed can be straightforwardly adapted to support OQoS priority-encoded packet transmission. The packet encoding mechanism is offered through a simple modification of the PSE's electronic routing control logic, in addition to the fabric's supported optical message format. According to the high or low class of service assigned to the packet, the corresponding priority class is encoded directly in the optical header signals.

In this implementation, a low-duty electronic pseudo-clock signal is experimentally distributed among all the PSEs. The clock consists of two short pulses per timeslot (*i.e.* per message duration). The frame and address header signals are experimentally modified to incorporate a one-bit priority (Figure 4.6). The implemented routing control is based on sequential electronic logic that samples the frame on the two pulses of the pseudo-clock. The first sampled bit determines the presence of the optical packet, while the second bit denotes the packet's OQoS class. In the case of a low-priority/OQoS packet, both of the sampled bit signals will be high. For a high-priority/OQoS packet, the first sampled bit will be high and the second bit will be low. The subsequent message routing decision at each PSE can then be made according to the two high/low levels detected by the control logic. In the situation of contention between two wavelength-striped packets, the new adapted routing logic and circuitry gates on the SOA associated with the high OQoS/priority packet, while dropping the contending low OQoS/priority packet.

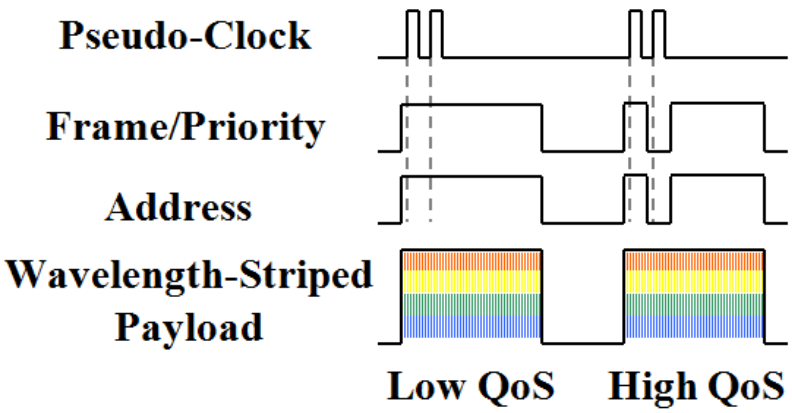

Figure 4.6: QoS Based Encoding Scheme - Block diagram depicting the implemented OQoS encoding scheme for the supported wavelength-striped optical packet.

4.2.2 Experimental Results

Figure 4.1 depicts the implemented switching fabric hardware that was used in this experimental demonstration. The electronic control logic for the OQoS encoding scheme is synthesized in the CPLDs within the PSEs, alongside the basic packet routing logic. A one-bit, two-level OQoS routing is implemented here as an initial demonstration of the feasibility of this approach; by using a pseudo-clock with multiple pulses, a multi-level QoS implementation could also be achieved.

To experimentally demonstrate the OQoS-encoded packet routing, a set pattern of wavelength-striped packets is injected in the OPS fabric with a combination of high and low priority encoded packets. This demonstration supports 57.6-ns timeslots, containing 51.2-ns duration packets with data modulated at 10 Gb/s on eight payload wavelengths (ranging from 1540.1 nm to 1558.3 nm). The pseudo-clock thus uses two pulses within the 51.2-ns packet duration to sample the packet's OQoS level. Optical packets are created with a 2^7-1 NRZ-OOK PRBS data stream and gated into packets

using external SOAs. The optical header of each packet has a frame signal encoding a one-bit OQoS, one-bit distribution address (selecting one of two possible paths through the fabric test-bed), and two-bit routing address. Figure 4.7 provides the input and output optical waveform traces of the pseudo-clock and optical packets, and confirms correct routing. The figure shows the packets' frame with encoded priority, address, and one sample wavelength channel of the 8×10-Gb/s payload. The faded waveforms in Figure 4.7 refer to the contending low-OQoS packets that are initially dropped due to the control logic's encoding scheme.

The experimental packet sequence exemplifies the functionality of the OQoS encoding technique. This exploration shows one high-OQoS source (in1), one low-OQoS source (in2), and one source whose retransmitted packets are given higher OQoS (in0) (these packets can be seen in Figure 4.7). When two messages contend at a given PSE, the lower-priority packet is dropped and no ack pulse is received at its sending port. If the sending port does not obtain an ack, it can retransmit on a different path in the next timeslot by modifying the distribution bit. The retransmitted packet can have an equivalent or higher priority class. Here, the acks are not implemented due to the large round-trip time of the realized test-bed compared to the envisioned integrated one; instead, packets are assumed to be received (or not) within the timeslot.

Error-free transmission of all egressing packets is verified with a 10-Gb/s direct-coupled *p-i-n* receiver with TIA and LA. The BERT is synchronized with the packet gating signals and is gated over 80% of the packet. BERs less than 10^{-12} are obtained for each of the eight payload wavelengths. Using a VOA before the receiver, sensitivity curves are recorded for one typical payload channel; these are shown in Figure 4.8, with

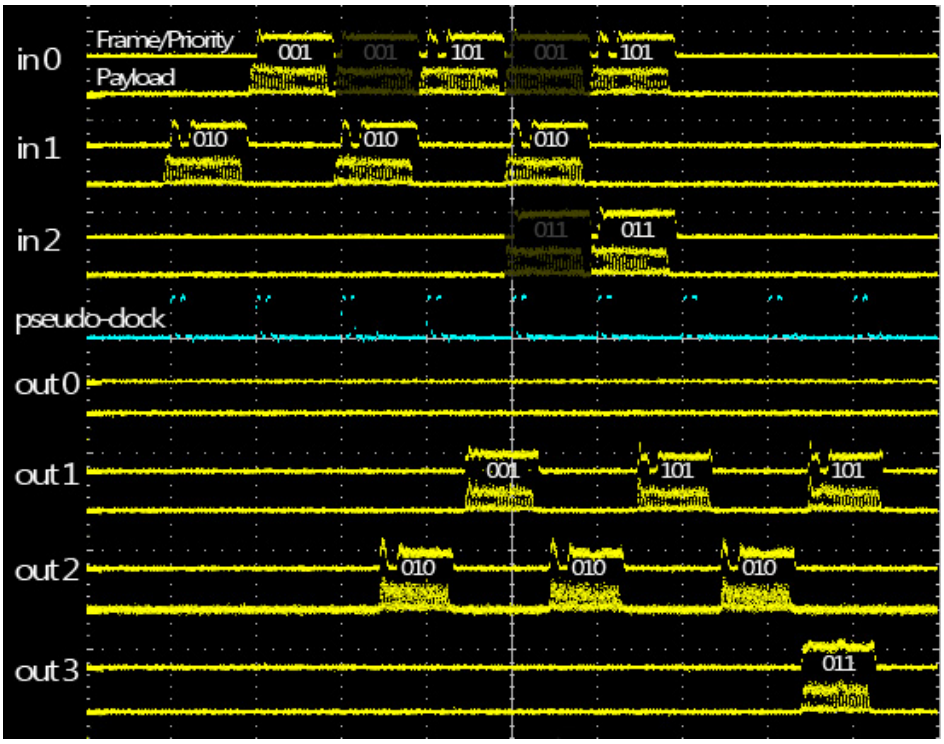

Figure 4.7: QoS Based Routing Traces - Experimental optical input and output waveform traces associated with the optical QoS based traffic, validating the correct routing of the encoding scheme. The pseudo-clock trace is also provided. Packets are injected using three input ports, with the three address bits labeled above the waveforms, and emerge from four output ports.

the corresponding 10-Gb/s input and output eye diagrams. All payload wavelength channels performed similarly. The three-stage test-bed exhibits a power penalty of 1.2 dB (equivalent to 0.4 dB per SOA hop), which is consistent with previously shown power penalty values at 10 Gb/s. The performance of the test-bed is not negatively affected by the realization of the encoding scheme. This verifies the functionality of the proposed OQoS-encoding routing scheme.

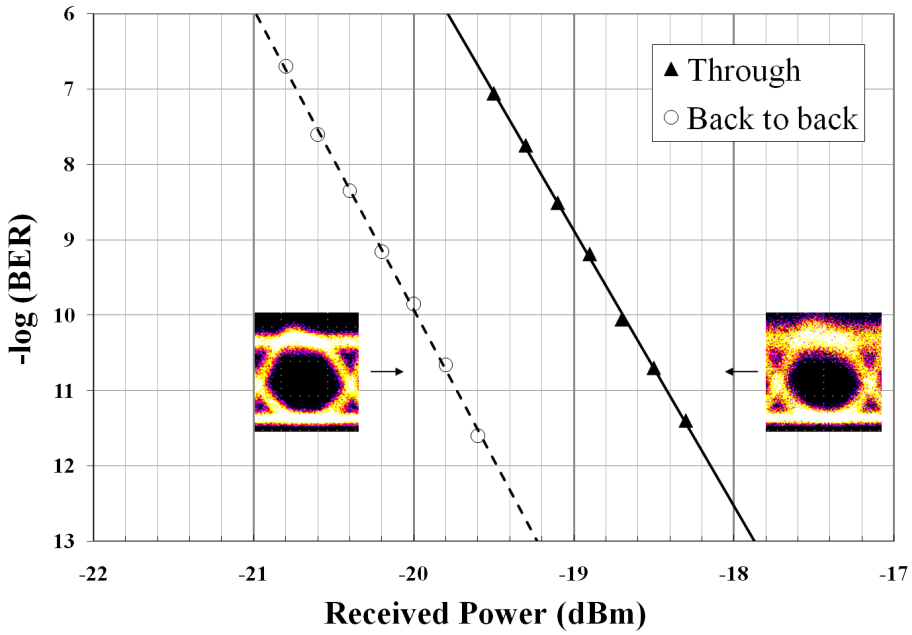

Figure 4.8: QoS Based Routing Sensitivity Curves - 10-Gb/s BER sensitivity curves for one typical payload channel (dashed line refers to the back-to-back measurements, solid line refers to the output data). Insets present the 10-Gb/s eye diagrams of the input and output (λ = 1558.28 nm).

Future packet routing applications will likely provide a prerequisite allowing for the priority of end users to be taken into account to ensure sufficiently high QoS for users with higher priority. The physical-layer switching fabric should thus be designed

to support transmission priority of high-QoS packets and data paths. In this work, the functionality of the optical layer is enhanced by implementing different optical classes directly on the physical layer. The priority encoding mechanism provides an OQoS encoding technique for routing wavelength-striped optical messages, effectively demonstrating two distinct classes of frame-encoded packet priority. The scheme offers high and low priority levels, as well as prioritized routing in the case of message-dropping. This exploration examines the potential and reaffirms the feasibility of realizing OQoS-aware protocols in the future Internet infrastructure.

4.3 Multi-Terabit Capacity

As this thesis has highlighted, the major challenge of delivering high-bandwidth, broadband user traffic is driving the development and deployment of optical systems. By leveraging emerging photonic technologies and optical network elements, increased bandwidths can be achieved by future optical routers. To this end, the optical switching fabric within these routers will need to support line-card rates at extremely high data rates, possibly with various advanced modulation formats [3, 4]. Assuming a wavelength-striped packet format, each payload WDM channel in the optical message may be required to be modulated at rates of 40 Gb/s and beyond. By supporting these high-speed network links, OPS networks can potentially achieve highly dynamic, intelligent, and programmable packet switching on the optical layer.

This work addresses this notion by showcasing the multi-terabit capacity of the implemented 4×4 optical switching fabric, that can seamlessly support the successful and error-free routing and transmission of wavelength-striped optical packets with

8×40-Gb/s payloads [26]. This yields an aggregate bandwidth of 320 Gb/s per network port. The complete switching fabric thus supports over a terabit of total optical bandwidth. BERs less than 10^{-12} can be achieved on all eight of the 40-Gb/s payload wavelength channels, with an average power penalty of 0.5 dB per SOA hop, taken at 40 Gb/s and at a BER of 10^{-9}. This demonstration illustrates the potential for achieving high-speed optical fabric lightpaths with error-free performance of wavelength-striped optical messages.

4.3.1 Multi-Terabit Transmission Experimental Demonstration

The transparent optical switching fabric design used here (Figure 4.9) is composed of four independent 2×2 PSEs, arranged in two stages with two PSEs per stage, providing a means of achieving high-bandwidth transparent optical lightpaths and interconnections for next-generation Internet routers and switches. The implementation is similar to other experiments, with the exception that the wavelength-striped packet uses 40-Gb/s modulated data rates on each payload segment. Previously, each payload channel was modulated at 10 Gb/s [25, 97]. In this demonstration, the fabric's broadband transparency, inherent with the use of SOAs, supports data rates of 40 Gb/s per payload wavelength channel. This demonstration uses eight wavelengths each at 40 Gb/s; however, this does not begin to approach the achievable limit. Higher data rates – per channel – are feasible, and planned future work (to be performed with the aid of the author) will truly leverage the fabric's transparency to show the support of various modulation formats. The test-bed's multi-terabit capacity is attested to

by the fact that multiple high-data-rate packets can be injected in all input ports of the fabric. Synchronous transmission of fixed-length packets is shown here, though asynchronous transmission may also be supported (as per [91]).

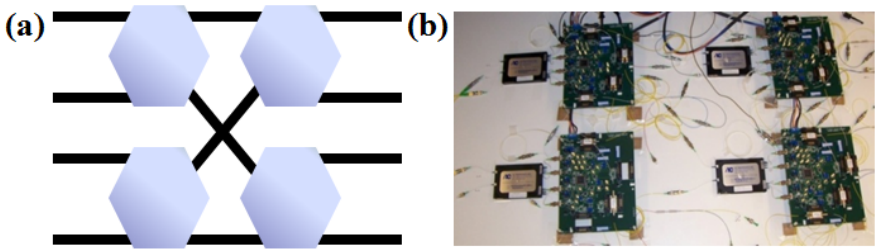

Figure 4.9: Multi-Terabit Capacity Fabric - (a) Block diagram showing the topology for the two-stage 4× switching fabric; (b) Photograph of the implemented fabric test-bed.

The test-bed here supports 8×40-Gb/s wavelength-striped optical messages, with 40-Gb/s OOK data encoded on eight payload wavelength channels. As seen in Figure 4.10, the experimental setup uses eight CW-DFB lasers ranging from 1540.56 nm to 1560.61 nm, which are each sent through a PC, then combined using a wavelength-division-multiplexer. All channels are simultaneously modulated with a single 40-Gb/s $LiNbO_3$ amplitude modulator, encoding a 2^{15}-1 PRBS on all eight payload channels. The 40-Gb/s RF signal that drives the modulator is created using a 10-Gb/s PPG and high-speed electrical multiplexer. The payload channels are then decorrelated using a span of SMF-28. The control wavelengths are generated independently, using separate CW-DFB lasers at desired wavelengths; namely, the control channels include the frame bit at 1555.75 nm (C27), as well as a two-bit address at 1552.52 nm (C31) and 1531.12 nm (C58). The data and control information are gated into 64-ns long optical packets using external SOAs driven by a DTG and combined using a passive coupler, resulting

in wavelength-striped optical packets with three control bits multiplexed with eight payload wavelength channels.

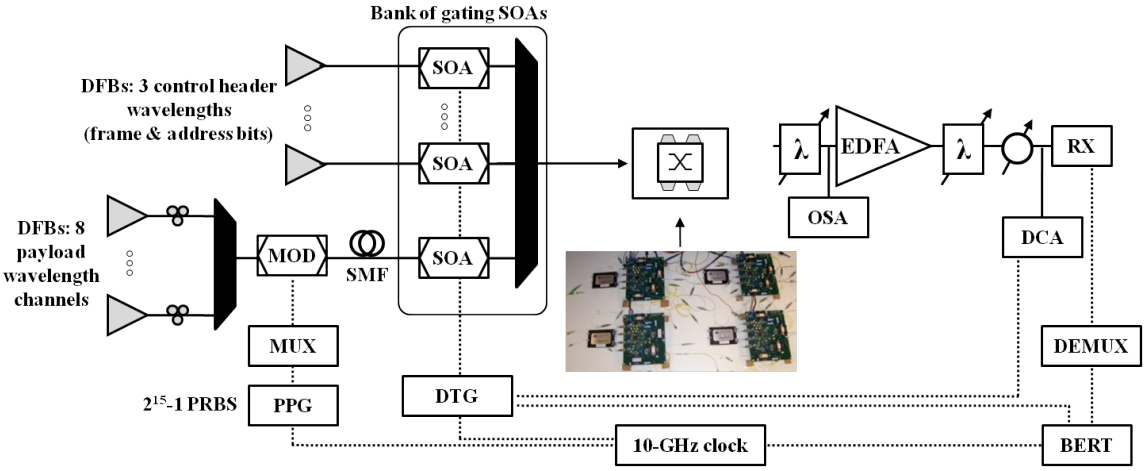

Figure 4.10: Multi-Terabit Experimental Setup - Diagram of the complete experimental setup.

The WDM optical packets are injected in the OPS fabric test-bed and switched using the SOA-based PSEs. The SOAs provide sufficient amplification to compensate for the passive losses incurred by propagation through the PSEs. Correct routing of the 8×40-Gb/s packets through the fabric is verified. At the output of the fabric, the packet analysis system involves transmitting the egressing packets to a tunable filter (λ in Figure 4.10) to select one 40-Gb/s payload channel and an EDFA. A second tunable filter is used to suppress the EDFA's ASE, and then transmits the signal to a variable optical attenuator, and subsequently to a high-speed 40-Gb/s receiver composed of a *p-i-n* photodiode and TIA. The received electronic data signal is then time-demultiplexed for evaluation by a 10-Gb/s BERT. The DTG is also used to gate the BERT on the optical packets. A common clock synchronizes the DTG, PPG, BERT, and electrical

multiplexer/demultiplexer, and no clock recovery is implemented here.

4.3.2 Multi-Terabit Transmission Results

The switching fabric test-bed is shown to correctly switch the 8×40-Gb/s wavelength-striped optical packets. BER measurements show that error-free transmission is achieved, obtaining BERs less than 10^{-12} on all eight 40-Gb/s payload wavelength channels. Figure 4.11a depicts the input and output optical waveforms of the packets with the encoded 40-Gb/s data. The 40-Gb/s input and output eye diagrams corresponding to four of the eight payload wavelengths channels is given in Figure 4.11b-e; namely, the 40-Gb/s eyes are shown for 1540.56 nm (C46), 1546.92 nm (C38), 1550.92 nm (C33), and 1558.98 nm (C23). No extinction ratio differences are seen as the packets propagate through the fabric test-bed.

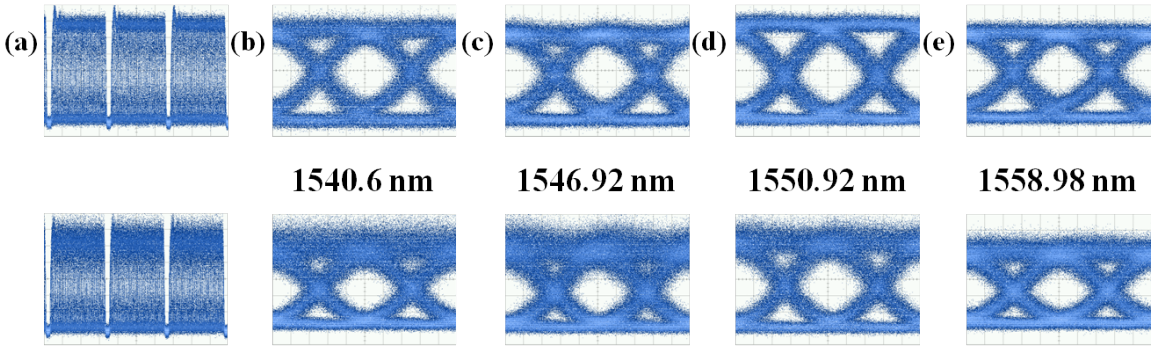

Figure 4.11: Multi-Terabit Waveforms and Eye Diagrams - (a) Input (top) and output (bottom) optical waveform traces of the optical packets; (b)–(e) 40-Gb/s input (top) and output (bottom) eye diagrams of a subset of the eight payload wavelength channels.

40-Gb/s sensitivity curves for one representative error-free channel (Figure 4.12) show a power penalty of 1 dB for the two-stage fabric, taken at a BER of 10^{-9}. For

this demonstrated transmission of wavelength-striped packets with 40-Gb/s data on each payload channel, the resulting average power penalty is 0.5 dB per SOA gate hop. The power penalty for 40-Gb/s data rates is higher than that at 10 Gb/s, which may comprise an important design factor for future large-scale network implementations.

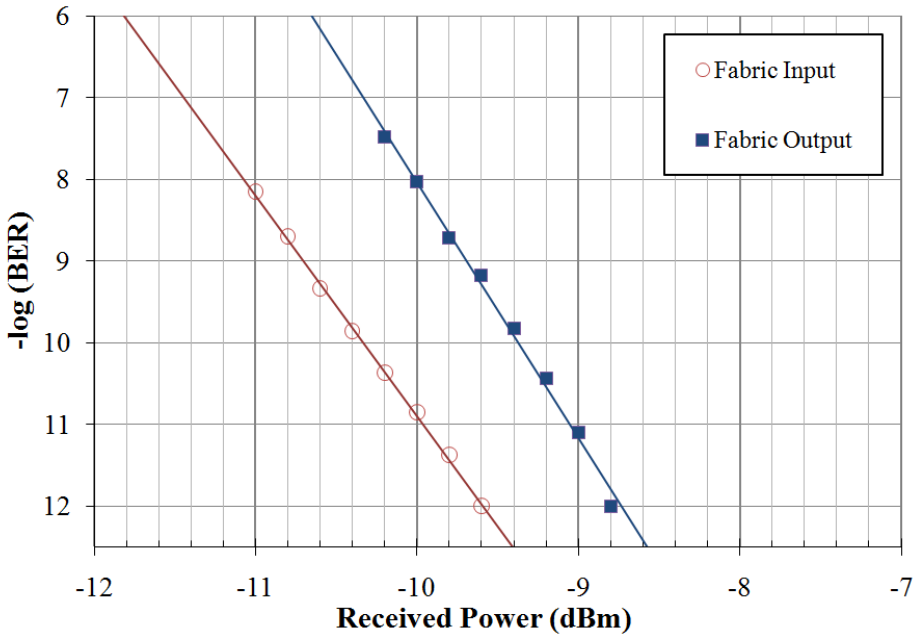

Figure 4.12: Multi-Terabit Capacity Sensitivity Curves - 40-Gb/s BER sensitivity curves for one typical payload channel (red line refers to the back-to-back measurements taken at the fabric input, blue line refers to the data taken at the fabric output) ($\lambda = 1550.92$ nm).

In short, the links in future optical networks and routers will be required to operate at very high data rates in order to accommodate the exploding demand in bandwidth. This experiment demonstrates an optical packet switching fabric that successfully supports wavelength-striped messages with 40-Gb/s data encoded on multiple payload wavelengths, achieving an increased aggregate bandwidth of 320 Gb/s per network port.

The implemented 4×4 fabric thus realizes over a terabit of optical switching capacity. This demonstration of high-speed optical packet routing and transmission constitutes a fundamental step in realizing the data rates required by future applications, specifically at 40 Gb/s and beyond. By leveraging the transparency of the fabric, one can imagine scaling the per-payload-channel rates to high data rates and/or other (advanced) modulation formats.

4.4 Packet Multicasting: An Overview

The success of the future cross-layer design will capitalize on the greater functionality envisioned on the physical layer. The following two sections discuss an extremely important high-level network functionality that can be realized at the optical switching fabric layer, namely *packet multicasting*. This effort highlights two possible architectures that can enable the multicasting capability on the switching fabric test-bed [119]. The application provides advanced control over multiwavelength optical messages, ideally with a packet-level granularity.

Optical packet multicasting is a remarkable exemplary application that enables greater functionality and programmable flexibility for future OPS fabrics [120]. Multicasting is an inherent characteristic of the IP layer that allows a single source to simultaneously transmit data packets to multiple destination endpoints. Multicasting drives high-bandwidth user-demand applications such as distributed/cloud computing, streaming of HD video images, networked gaming, interactive online conferencing, and forward-looking telemedicine. By migrating this functionality lower in the OSI stack to the optical layer, these broadband packet-based applications can be envisioned to be

supported directly on the underlying optical network, effectively with lower cost [39].

Given that the energy consumption associated with telecommunication networks is predicted to grow exponentially in the next decade [70, 72], packet multicasting may be a feasible approach to lower the network's operating energy costs. Redundant O/E/O conversions may be avoided on the lower network layers, mitigating energy-costly IP routing, as confirmed by Tucker *et al.* [77]. Optical packet multicasting will allow for the simultaneous transmission of multiple packets, avoiding expensive retransmissions by the IP-layer routers, and thus comprising a way to minimize the total energy consumption associated with IP-layer transmission and routing. Figure 4.13 shows the envisioned cross-layer stack with how the optical multicast operation may be integrated.

Here, the specific focus is on implementing broadband packet (or waveband) multicasting for OPS fabrics, wherein wavelength-striped optical messages can be transmitted from a single source input port to a subset of the destination output ports. Previous analytical investigations on the multicast scheduling problem assume an N-input, N-output port packet switching architecture that is intrinsically capable of multicasting and broadcasting [121, 122]. These studies provide packet switch analyses and represent both stochastic and deterministic performance evaluations, concluding that the issue of multicast scheduling is NP-hard. However, there has been very little previous work demonstrating the optical multicast operation in an experimental test-bed environment. Furthermore, prior multicast-capable OPS designs focus on wavelength multicasting [123, 124] that require wavelength converters, or use impractical optical buffers or FDLs [125, 126] to realize the multicasting capabilities.

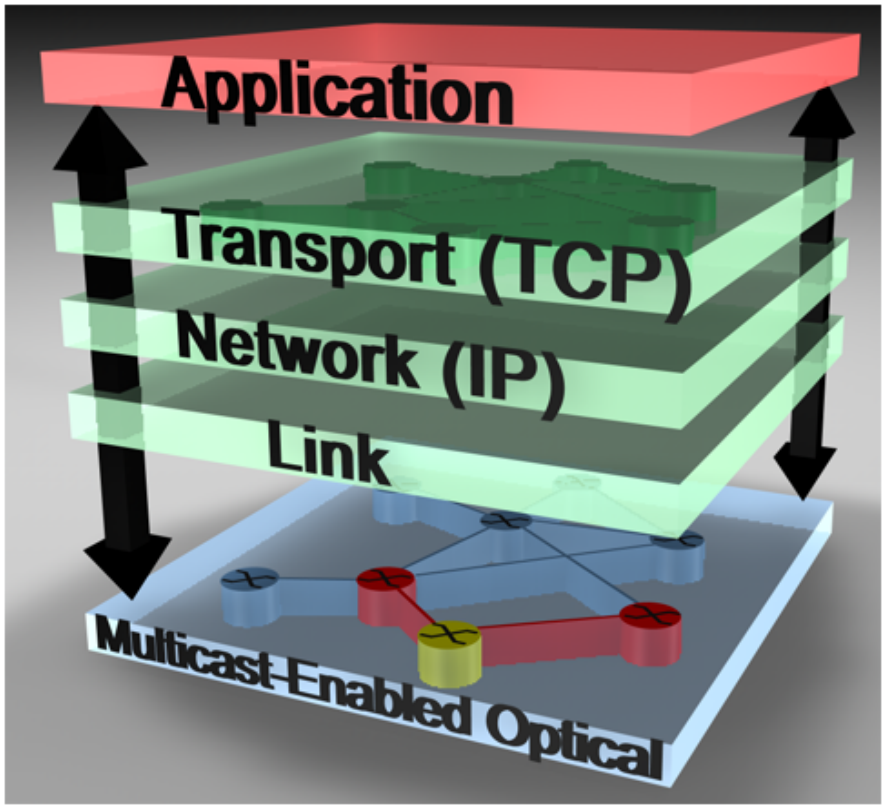

Figure 4.13: Packet Multicasting Vision - Illustration of a future cross-layer optimized network stack, with bidirectional signaling between the network layers. The optical physical layer can provide an integrated optical packet multicast operation, where one network node (yellow) can simultaneously transmit to multiple nodes (red).

Here, the bufferless OPS fabric implements the multicasting of wavelength-striped packets and thus avoids the need for wavelength converters or optical buffers. This will facilitate supporting the necessary high bandwidths, since the multiwavelength packet can easily support per-channel data rates of 10 Gb/s, 40 Gb/s, or higher, as required by future broadband user applications [22, 26].

This work takes advantage of the optical switching fabric architecture's distributed electronic routing logic control to seamlessly support the multiwavelength packet multicast operation in an optical test-bed. Two distinct optical switching fabric architectures are investigated and the two packet multicast-capable fabric designs are experimentally implemented on a programmable 4×4 OPS fabric in the test-bed. Figure 4.14 depicts the envisioned future optical-layer structure, showing how the two multicast-capable switching fabric designs could be incorporated. The first design is based on a splitter-and-delivery (SaD) architecture, first proposed in [127] as a means to realize wavelength multicasting. Here, the original SaD design is modified to enable a programmable wavelength-striped packet-splitter-and-delivery (PSaD) functionality, to provide a higher level of connectivity and support the multicast of multiwavelength messages [128]. The WDM packet is switched through the fabric without using wavelength conversion techniques. Two parallel OPS switches are realized wherein each supports a unicast operation to yield a non-blocking two-way multicast through the fabric. This first experiment shows the error-free multicasting of multiwavelength optical messages to multiple destination ports, transmitting 8×10-Gb/s wavelength-striped payloads with BERs less than 10^{-12}. Scalability of the per-channel data rates is confirmed to 40 Gb/s.

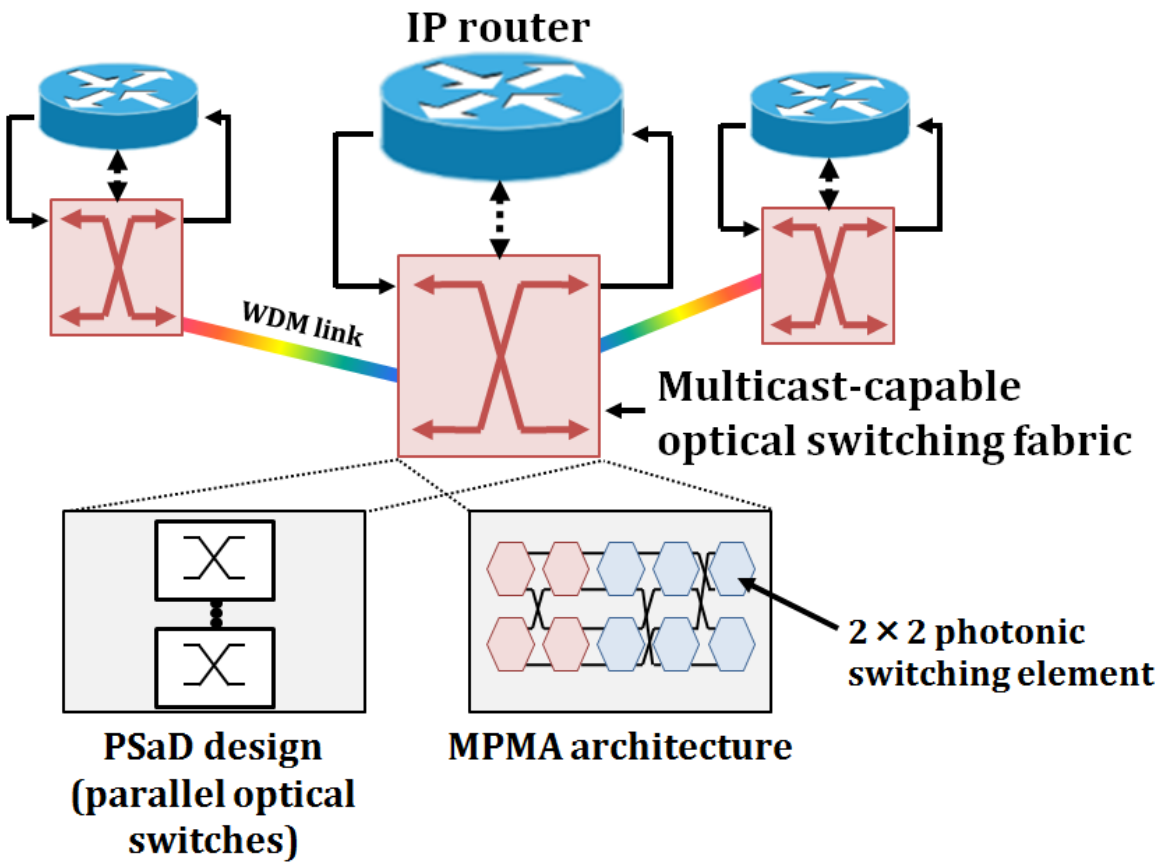

Figure 4.14: Network Architecture with Packet Multicasting - Diagram of the envisioned network node, depicting how the multicast-capable optical switching fabric designs would be integrated with IP routers. Two possible multicasting architectures are discussed: a PSaD design, comprised of multiple internal optical switches; and the proposed MPMA architecture. Both switching fabric designs use building blocks consisting of 2×2 photonic switching elements (PSEs).

In the following sections of this dissertation, the Multistage Packet Multicasting Architecture (MPMA) is then introduced, showcasing an improved network design for broadband multicasting that has a lower hardware cost [129] to implement a higher-radix multicast. MPMA employs a multistage design that uniquely leverages the programmable nature of the switching fabric architecture and is implemented in the experimental test-bed, capitalizing on the reprogrammable 2×2 optical switching elements. The error-free routing and multicasting of 8×10-Gb/s optical packets is achieved with BERs less than 10^{-12} and a 2.5-dB average power penalty for the 5-stage network, taken at a BER of 10^{-9}. For both distinct experimental demonstrations, the switching fabric architectures are shown to support the simultaneous transmission of multiwavelength optical packets in a programmable fashion, as well as the seamless support of the unicast, multicast, and broadcast operations. The two distinct designs showcase the design trade-offs that exist between multicast routing complexity and network hardware costs.

4.5 Packet-Splitter-and-Delivery Multicasting Design

The first multicast-capable architecture discussed in this work uses a programmable packet-splitter-and-delivery design that can support the simultaneous transmission of multiple broadband, wavelength-striped messages to multiple outputs [128]. The initial basic SaD architecture is non-blocking and splits the incoming lightpath into M spatial outputs to deliver the lightpath to M separate destination endpoints. Here, the original architecture is modified to create a PSaD system with a higher level of

connectivity, where the input wavelength-striped packet can be split multiple ways to enable the multicasting capability. This is then similar to the splitter-and-combiner structure discussed in [130]; however, since this system supports multiwavelength packet multicasting, it does not require the wavelength conversion subsystem.

The proposed design leverages an optical switching fabric that is internally composed of M parallel optical packet switches interconnecting N network terminals (Figure 4.15). In order to realize the multicasting functionality, each source input is connected to each destination output using M separate switch entities that operate in parallel. The PSaD architecture creates M distinct and independent paths between each source and destination, in a non-blocking fashion. Each path (*i.e.* optical switch) supports the multiwavelength optical packet format. One clear advantage of this design is that the optical switching fabric can either handle a unicast using a single switch, or multicast using combinations of several of the switches.

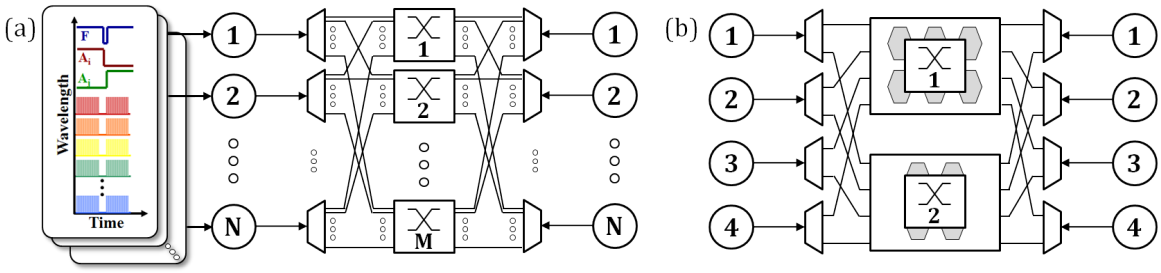

Figure 4.15: Multicast-Capable PSaD Architecture - (a) Proposed packet-splitter-and-delivery (PSaD) architecture that supports wavelength-striped optical messages ingressing on N input ports, using M optical packet switches. The design uses optical packet switches that operate in parallel; (b) Block diagram of PSaD design demonstrated in the experiment, supporting N=4 input ports with M=2 internal switches. The optical packets pass through optical splitters and combiners at the input and output of the fabric, respectively.

For the purposes of the fabric's test-bed demonstration, a two-way multicasting is realized using a complete 4×4 optical switching fabric that is comprised of two OPS switches placed in parallel (*i.e. N=4* and *M=2*) (Figure 4.15b). The two internal switches provide two independent paths in order to realize a multicast to two distinct destinations through the fabric. The incoming multiwavelength packet is injected into the switching fabric in the test-bed via propagation on a single fiber. The optical packet is split using a 1:2 passive optical coupler to create two replicas of the multiwavelength packet, each of which continues to propagate to the respective input ports in the first routing stages of the two switches. At the output of the fabric, another 2:1 passive optical coupler is used to combine the packets egressing from both switches from the corresponding output ports. In future implementations with $M>2$ (*i.e.* where the number of parallel optical switches is greater than 2), a $1\times M$ SOA-based switch can be used to provide gain to compensate for the insertion loss of splitting the signal M ways at the input (in place of the optical couplers). Similarly, a $M\times 1$ SOA-based switch can be deployed as a combiner at the output.

Here, the upper parallel switch is based on a three-stage banyan architecture, organized in an enhanced Omega interconnection network topology, with a distribution routing stage [25], using a total of six 2×2 PSEs. The lower parallel switch is based on a two-stage banyan topology, using four 2×2 elements. Thus, using the three-stage switch in parallel with the two-stage switch, the complete PSaD design is experimentally implemented using ten 2×2 PSEs with five routing stages, with two PSEs in each stage. Each of the distributed switches uniquely supports a unicast; thus, by operating several of these switches in parallel, the complete architecture supports the desired

packet multicasting operation. Further, it should be noted that this implementation interconnects *N=4* input and output ports using ten PSEs, which is not the absolute minimum; a similar connectivity for a *M=2* fan-out multicast could be achieved using eight PSEs (using two two-stage switches in parallel). In this realization, the flexibility of the switching fabric is leveraged by demonstrating that varying switch topologies could be deployed in the M parallel switches with little effect on the overall multicasting functionality.

Here, as the leading edges of the two optical packets reach the first stage's PSEs, the messages are switched according to the optical headers encoded in the packet. For the five total routing stages, the optical packet leverages five distinct routing address bits. According to these extracted address bits, the two optical packets are routed through the two implemented OPS switches. Within each of the two switches that make up the complete switching fabric, the packet multicasting operation is realized since each switch supports the high-bandwidth wavelength-striped optical packets, which are switched entirely in the optical domain. The wideband nature of the SOA-based PSEs allows for a straightforward realization of spatial multicasting of packets composed of multiple wavelengths. The multicasted multiwavelength optical packets are delivered to their desired (multiple) destinations at the output of the complete switching fabric, as per the required multicasting request encoded in the headers. The injected optical packet can be unicasted on a single switch (either upper or lower), or multicasted by traversing both entities simultaneously to reach multiple, distinct output destinations.

The packet multicasting functionality leverages the unique programmability of the individual PSEs. By allowing the PSEs to act as identical building blocks that can be

arranged in different topologies, the multicasting operation can be straightforwardly realized by placing multiple of the switch entities in parallel. Furthermore, taking into account the distributed nature of the PSEs' routing logic and the fact that there is no signaling required between PSEs, or between the PSEs and a centralized control plane, implementing several of the OPS switches in parallel is a scalable way to enable packet multicasting. The number of realizable switches does not grow with a required controller unit or logic. The simple architecture shows multicasting with limited added routing complexity. Within this design, optical buffers and wavelength converters are not required to support the packet multicasting functionality. One may argue that to achieve a M-way multicasting, one would require M parallel optical switches, which could potentially be costly to implement. The second MPMA design presents a possible solution to this issue.

4.5.1 Experimental Demonstration and Results

The first experimental demonstration of the multicast-capable PSaD design shows the correct routing of multiwavelength optical packets incorporating 8×10-Gb/s wavelength-striped payloads. The packets are multicasted error-free to multiple destination ports using an implemented optical switching fabric with BERs confirmed less than 10^{-12}. A pattern of wavelength-striped optical packets is injected in the fabric, and thus simultaneously into both optical switches (Figure 4.16). The test-bed connects four independent input ports to four distinct output ports using the two parallel switch entities, thus offering a path diversity of two separate routes from a given input to a given output. The payload data for the wavelength-striped packets

are generated using eight CW-DFB lasers ranging from 1533.18 nm to 1564.39 nm, which are combined onto a single fiber using a passive 8:1 optical multiplexer. The eight wavelength channels are then simultaneously modulated with a 10-Gb/s NRZ-OOK signal that carries a 2^7-1 PRBS. The wavelength channels are decorrelated by 25 km of SMF-28. The payload wavelengths are then split using a passive 1:3 optical coupler to create three modulated wavelength-striped data flows for injection in three fabric ports. Each set of payload wavelength signals are then transmitted to external gating SOAs.

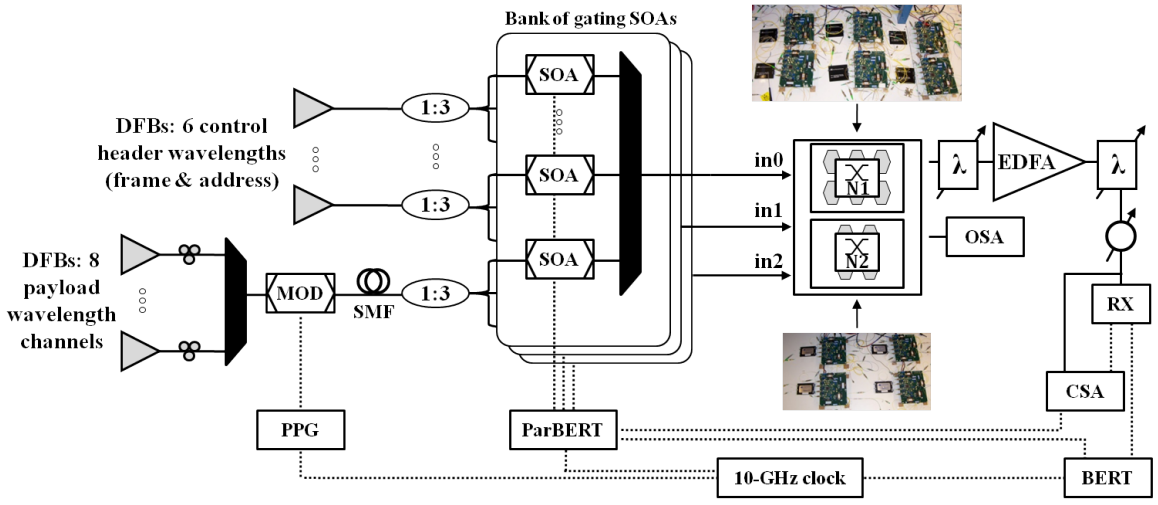

Figure 4.16: Multicast-Capable PSaD Experimental Setup - Setup for the PSaD demonstration, with photographs of the two switches internal to the optical switching fabric.

The six control wavelengths are generated using separate CW-DFB lasers, including one frame at 1555.75 nm and five address bits, ranging from 1531.12 nm to 1550.92 nm. The DFB lasers are split using 1:3 couplers and sent to a set of gating SOAs. The control header and payload data signals are then gated into packets using an array

of gating SOAs, encoding the appropriate addressing information for each packet to be routed through the test-bed. The control headers and the payload signals are then combined using a passive optical coupler to create the multiwavelength data stream. A similar packet-generation setup is used in parallel for each collection of control and payload signals to form three distinct packet patterns for the three fabric input terminals. The ParBERT controls the gating SOAs for packet gating and fabric addressing. The ParBERT is pre-programmed with test packet patterns that are custom-designed for this demonstration. This system here thus creates 8×10-Gb/s wavelength-striped packets, with a six-wavelength control header and an eight-wavelength payload. These packets the ingress into the active ports of the switching fabric, simultaneously in both parallel OPS switches. The switch selection (*i.e.* whether packets are transmitted on the upper or lower optical switch within PSaD) is based on an *a priori* knowledge of the address wavelength space and the destination availability during the custom test pattern. The experiment here supports timeslots that are 128 ns in length, featuring optical packets with 115.2-ns durations. The header wavelengths for the switching and multicasting of each optical packet are predetermined for each experiment and programmable using the ParBERT.

At the fabric's output, the multiwavelength packet is monitored using an OSA and oscilloscope. The typical packet analysis system is also used in this experimental demonstration (Figure 4.16). The received electrical signals from the DC-coupled 10-Gb/s *p-i-n* photodiode with TIA and LA (RX) are sent to a BERT that is synchronized with the PPG and can be gated to analyze the optical packets with an electronic signal from the ParBERT.

In this way, the wavelength-striped optical packets are generated, injected in the implemented fabric, and routed through both parallel OPS switches. The messages are multicasted to two different destinations (if desired) by unicasting on each switch. The waveform traces associated with the optical packet traffic sequence in this experiment are given in Figure 4.17. The resulting packets egressing from the switching fabric test-bed are also given. The waveforms in Figure 4.17 provide the frame bit of the packet (set to high for the duration of the packet), as well as one of the payload wavelength channels, for each of the ingressing and egressing fabric ports. The two-bit binary addresses are indicated for each optical message in Figure 4.17. Since one address bit is required for each routing stage in the optical switch entity, packets routed through the three-stage N1 switch require three bits while messages routed using the two-stage N2 switch require two address bits.

All of the 8×10-Gb/s multiwavelength optical messages are correctly routed through the complete switching fabric, and accurately emerge at the destinations that are encoded in the control address headers. The multicasting operation is clearly validated, as the wavelength-striped packets are successfully routed from one fabric input port to multiple output ports.

The packet sequence shows that the switching fabric seamlessly supports both the unicast operation using a single switch entity, in addition to the multicasting operation with both switches. In the first active timeslot depicted in Figure 4.17 (denoted as A), the sources at two independent input ports transmit only on the upper fabric switch (N1, using in0 and in2). The optical packet from in0 has an encoded address of 001, which represents its output port (out1); the packet clearly emerges from this output

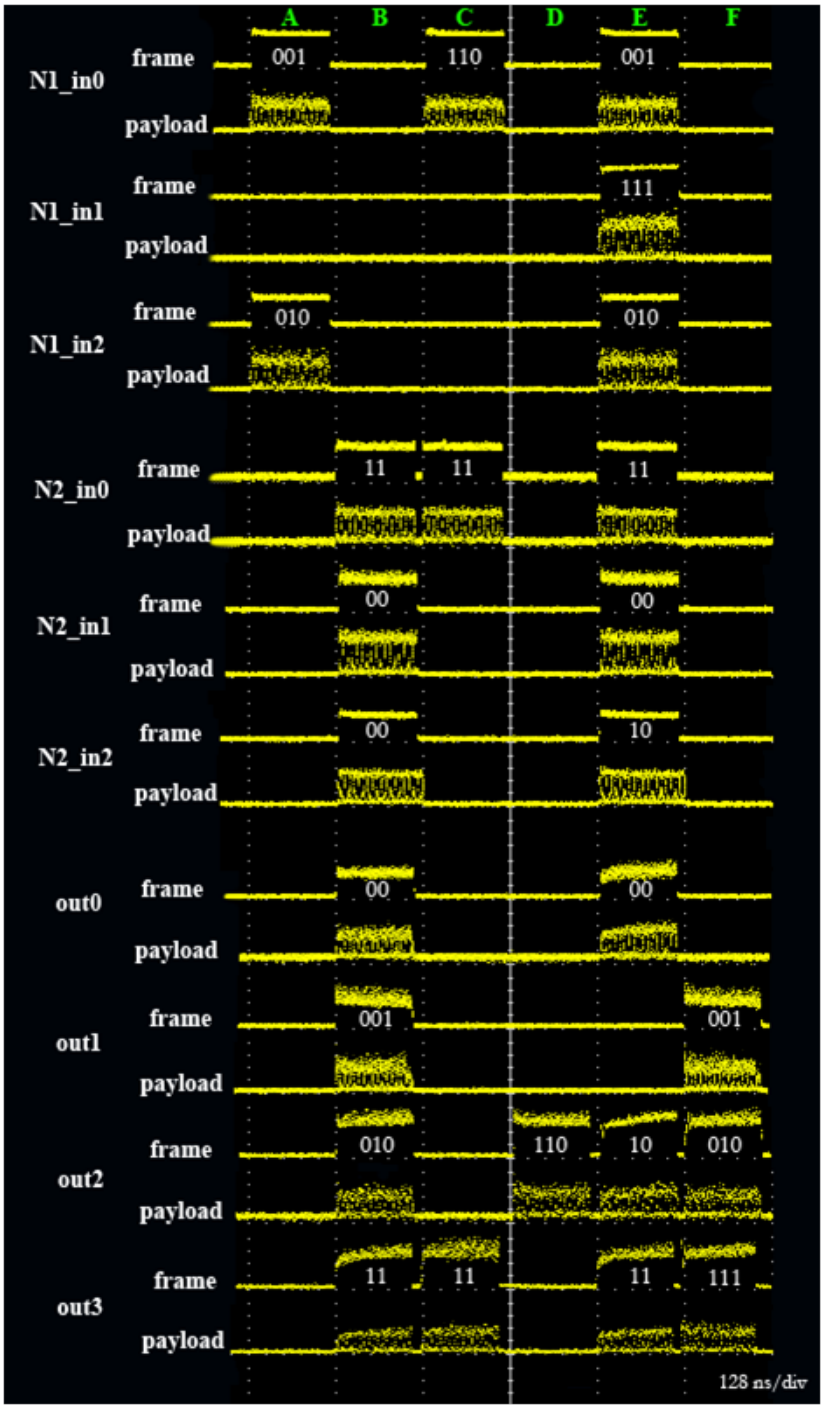

Figure 4.17: Multicast-Capable PSaD Waveforms - Experimental waveform traces of the input and output optical packet traffic patterns, with labels referring to the address information encoded in the optical packets.

port in timeslot B. Simultaneously, in timeslot A, a packet appears at the in2 port, with address 010 (addressed for output port out2), and the packet emerges from out2. Similarly, during the second active timeslot, all three sources transmit packets to three distinct output ports, each unicasting on the second/lower optical switch (N2). Packets have a two-bit address header, indicating their desired destinations: the packet from in0 (address 11) wishes to be routed to out3, the packet from in1 (address 00) has out0 as its required output, and the packet from in2 (address 00) has out0 as its desired destination. One of the contending packets from in1 and in2 are dropped and is retransmitted at a later timeslot. Note that due to the fact that the three-stage N1 has an additional routing stage as compared to the two-stage N2 (and thus larger transmission latency), packets that are routed through N2 appear one timeslot earlier than those routed through N1. Correspondingly, in the third timeslot C, a single source at the input of the switching fabric (in0) attempts to perform a packet multicast to two distinct destinations. This is done via two simultaneous unicast operations using both of the fabric's switches: packets are routed to out2 using N1 and to out3 using N2. These packets emerge from the output of the test-bed in timeslots D and E, respectively. During the fourth active timeslot for packet injection (timeslot E), all available sources attempt to multicast wavelength-striped optical messages to multiple output destination ports by simultaneously transmitting using both switching fabric entities, the upper N1 and the lower N2. These packets egress from the fabric test-bed after the appropriate delay. Packets routed through N2 appear at the output ports one timeslot earlier than those routed through N1, and all packets injected in timeslot E are correctly routed, appearing in timeslots E and F.

BER measurements confirm that all packets are received error-free, achieving BERs less than 10^{-12} on all eight payload wavelengths; this error-free transmission is verified after achieving zero errors in 10^{12} bits. Figure 4.18 shows the BER sensitivity curves corresponding to the back-to-back operation, as well as transmission through the lower N2, taken for the payload channel at λ=1541.05 nm. The insets show the optical eye diagrams for the same 10-Gb/s channel. An approximate 1-dB power penalty is measured for a two-SOA hop system (N2), corresponding to a 0.5-dB penalty performance for each SOA transversal.

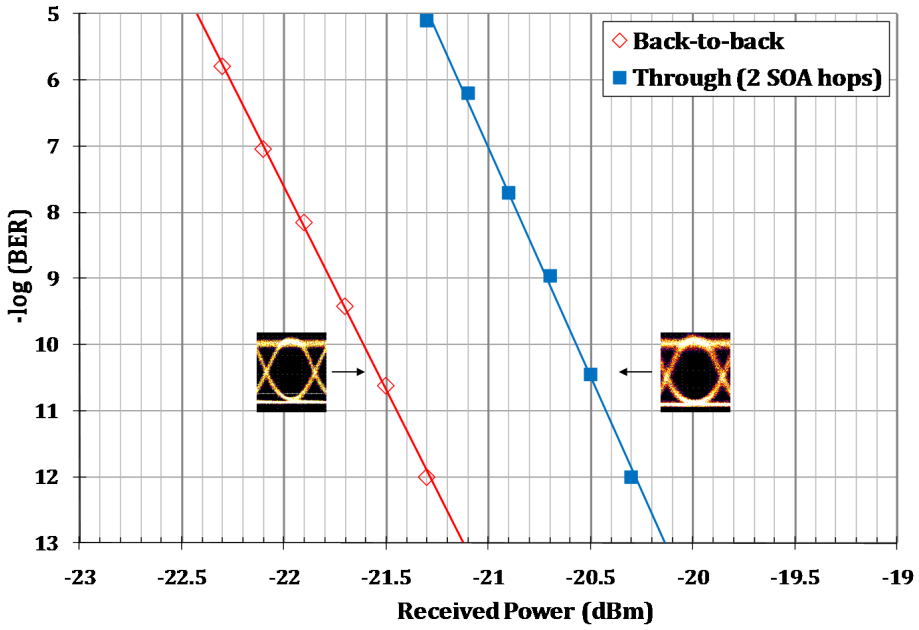

Figure 4.18: Multicast-Capable PSaD Sensitivity Curves - BER sensitivity curves corresponding to the PSaD demonstration, with insets showing the 10-Gb/s input and output eye diagrams associated with one payload wavelength channel (λ=1541.05 nm).

Additionally, the inherent bit-rate transparency of the switching fabric is capitalized

here to further confirm the scalability of the packet payload channels' data rates to higher modulation rates; the multicasting functionality is demonstrated using these high per-channel bit rates. The bit rates of the modulated data streams are scaled from the initial 10 Gb/s per channel to 40 Gb/s per payload wavelength. In this way, the switching fabric supports an aggregate packet bandwidth of 250 Gb/s, composed of six 40-Gb/s and one 10-Gb/s multiplexed channels. The 10-Gb/s channel is used to demonstrate the error-free performance of the fabric, using the 10-Gb/s BERT packet analysis system above.

The experimental setup for the coupled 10-Gb/s and 40-Gb/s demonstration is similar to the 8×10-Gb/s setup outlined previously. The packets are generated using six CW-DFB lasers, ranging from 1533.18 nm to 1564.39 nm, whose outputs are multiplexed onto a single fiber. The six wavelength channels are then modulated using a 40-Gb/s $LiNbO_3$ modulator with a 2^7-1 PRBS signal in a NRZ-OOK format. A single 10-Gb/s channel is also simultaneously generated using a separate CW-DFB laser and 10-Gb/s $LiNbO_3$ modulator. All modulated payload channels are then multiplexed together with the appropriate control header signals, and gated with an external SOA in a similar fashion as described above.

Correct routing is achieved with these high-bandwidth multiwavelength packets. Figure 4.19 provides the optical eye diagrams of one of the filtered 40-Gb/s data streams (at λ=1558.24 nm) for the back-to-back input packet that is injected into the test-bed (as seen directly after the gating SOA) and for the output packet (as observed directly after the three-stage upper N1 optical switch). Error-free performance with BERs less than 10^{-12} is obtained on the 10-Gb/s stream. BER packet analysis for the 40-Gb/s

payload channels was not feasible at the time of this demonstration due to experimental limitations. However, by leveraging the multiplexed combination of the 10-Gb/s and 40-Gb/s payload channels, the successful transmission and multicasting of 250-Gb/s aggregate bandwidth optical packets are shown.

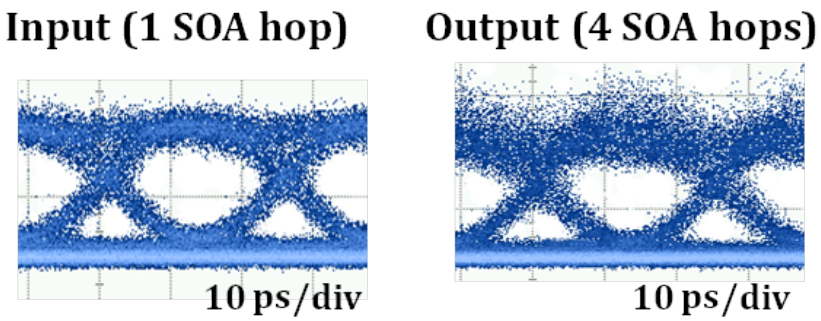

Figure 4.19: Multicast-Capable PSaD 40-Gb/s Eye Diagrams - 40-Gb/s eye diagrams verifying the feasibility to scaling the data rates of the individual payload wavelength channels beyond 10 Gb/s. The 40-Gb/s input eye (left) is taken directly after the gating SOA (before injection in the test-bed), while the output 40-Gb/s eye (right) is taken at the output of the switching fabric. The operating wavelength is λ=1558.24 nm).

4.5.2 Discussion

This packet multicast architecture leverages several parallel optical switches that operate in a distributed fashion. One can then design the complete switching fabric with differing topologies deployed in the internal switches that have various features such as a completely non-blocking design or one with low latency characteristics, among others. In this way, a traffic classifier may be realized at the source node to route optical messages according to the requirements of the transmitting applications. For example, optical packets that must be routed with minimal latency can be routed

through the fabric entity that supports this feature, or high-priority messages can be sent on the non-blocking switch to ensure successful transmission without the risk of packet contention.

4.6 Multistage Packet Multicasting Architecture

The previous PSaD design uses identical, simple routing logic within each PSE and achieves packet multicasting by simply adding replicated copies of a unicast-capable OPS switch in parallel. This may be costly to implement as the required multicast fan-out increases; therefore, this work now introduces a new architecture, MPMA. MPMA is an improved switching fabric topology for optical packet multicasting that is more efficient in terms of the additional hardware required [129].

With MPMA, the design of the switching fabric topology itself is optimized to minimize the added hardware costs. This architecture enables packet multicasting with a single optimized topology, requiring fewer additional components, and thus could potentially present cost savings in terms of minimizing energy consumption. Given that increasing network energy efficiency is a key driver [74, 77], the MPMA design may facilitate achieving an optical packet multicast that is cost-effective in future networks. MPMA capitalizes on the distributed nature of the PSEs' electronic routing logic and particularly on the unique high level of reprogrammability of the fabric's PSEs. By slightly increasing the complexity of the routing logic within the fabric's PSEs, a multicast architecture can be achieved that is significantly more hardware-cost efficient than the above PSaD design.

MPMA is an optimized multistage design that supports two distinct routing truth

tables in the PSEs. Both of the logic tables are based on simple combinational logic such that no centralized control is required for managing the PSEs, *i.e.* no signaling is required between the PSEs and a control unit to actuate the multicasting operation. Thus, the PSEs within the fabric do not all contain identical routing control logic. By allowing this small increase in routing complexity, this can enable a single switching fabric topology that is designed to use more routing stages to realize one-way, two-way, and four-way packet multicasting. Using the same amount of experimental hardware as the PSaD design, the achievable multicasting fan-out can be effectively increased to a four-way multicast. Similar to the PSaD realization, the design also easily supports unicasting, multicasting, and broadcasting functionalities. However, one trade-off is the decreased path diversity within the topology to manage contention. Messages that contend within the switching fabric are dropped and thus must be retransmitted in a subsequent timeslot.

The MPMA design is comprised of one subset of fabric stages that is used for packet routing (PaR), followed by one subset of stages that is used for packet multicasting (PaM). The PSEs in the routing and multicasting stages contain differing control logic as realized by the PSEs' electronic CPLD (Figure 4.20). The 2×2 PSE's logic determines whether the ingressing packet will be sent to one or both of the PSE's output ports. In the case of the PaR stages, a multiwavelength packet that reaches one of the two input ports is routed directly to one of the output ports; *i.e.* depending on the address bits decoded by the electronic logic and circuitry, the CPLD will gate one of the corresponding SOAs (or none if contention occurs). In the case of the PaM stages, depending on the recovered address bits, an incoming wavelength-striped optical packet

can be routed to either one or both of the PSE's available outputs. The CPLD will gate either one or two of the SOAs associated with the message. In both cases, the routing and multicasting operations depend solely on the optical packet headers that are extracted from the message using fixed wavelength filters and low-speed optical receivers. By cascading combinations of PaR and PaM stages, various multicasting topologies can be realized to enable various multicast fan-outs.

Here, one distinct MPMA design is implemented to create a 4×4 optical switching fabric. The goal is to create an architecture that allows any input port to transmit to any single output port (unicast, or one-way multicast), as well as to a subset of output ports (two-way or four-way multicast). It is also important to note that due to the nature of the implemented 2×2 PSE hardware, the PSEs here in a single routing stage extracts one address bit; additionally, the PSEs in each stage use the same wavelength addressing. These factors limit the possible MPMA designs that are feasible in this test-bed environment. Ideally, the MPMA design would use a PSE whose switching state can be set to implement an integrated PaR/PaM logic. In order to simultaneously support the unicast and multicast operations, a 4×4 switching fabric using the aforementioned PSE design requires a minimum of five stages, with the first two stages acting as PaR stages and the last three stages as PaM stages (Figure 4.20a). Two different routing decision logic tables are distributed among the ten required PSEs. The multicasting operation allows high-bandwidth lightpaths from a single input to fan out to several output destinations. The optical message's headers indicate whether the message is required to unicast, multicast, or broadcast through the single optimized fabric topology. If the PSE hardware could support extracting two distinct

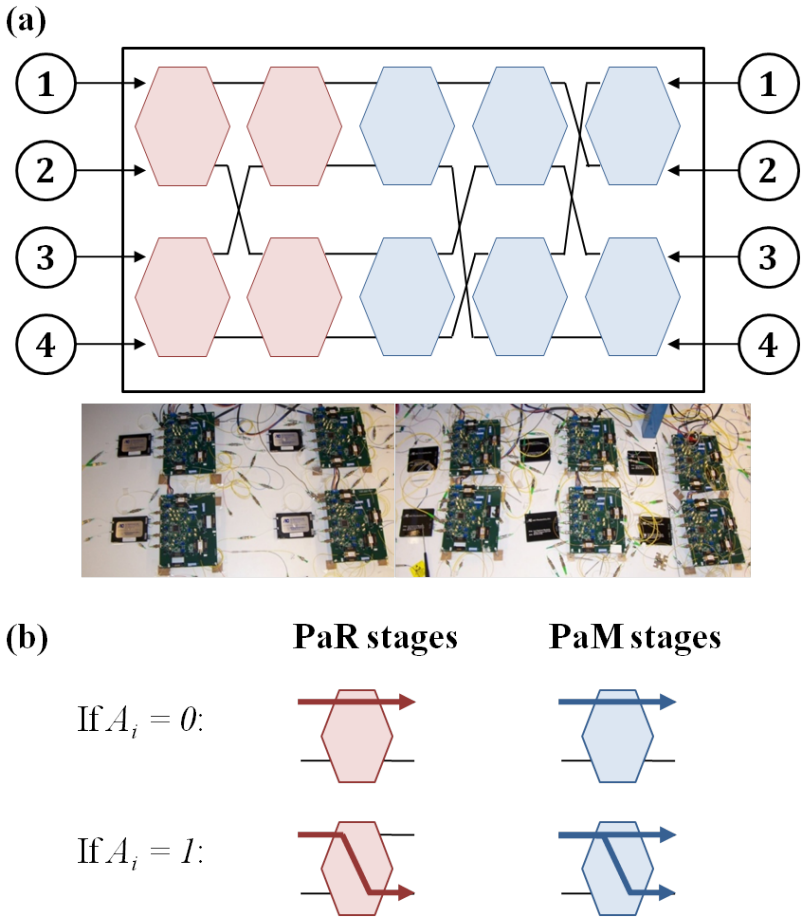

Figure 4.20: MPMA Architecture - (a) Block schematic of the proposed multicast-enabled fabric architecture, supporting four input/output ports with programmable PSEs implementing PaR (red) stages and PaM (blue) logic stages. A photograph of the implemented test-bed is shown below; (b) Block diagrams showing the routing logic in the PaR and PaM stages; depending on the low or high value of the stage's address bit, the multiwavelength packet is routed accordingly.

address bits at each stage (in addition to the frame bit), it would then be feasible to realize the 4×4 multicast-capable fabric design using fewer (*i.e.* 2 distinct) stages, in a broadcast-and-select topology, with a integrated version of PaR/PaM routing logic. This alternate topology would result in a more complex fabric addressing scheme as well as a more complicated PSE hardware design; however, fewer stages would be required (translating to fewer SOA hops), thereby improving the scalability of the multicast-capable topology.

4.6.1 Experimental Demonstration and Results

By simply experimentally reprogramming the control logic synthesized in the CPLDs of the ten 2×2 PSEs comprising the OPS fabric in the test-bed, the MPMA design can be straightforwardly implemented without additional hardware as compared to the PSaD design. As well, while a maximum fan-out of two is supported in the PSaD implementation, a maximum multicasting fan-out of four is achieved with MPMA with the identical hardware and equivalent number of components. The ability to multicast optical messages is shown by supporting the simultaneous transmission of 8×10-Gb/s wavelength-striped optical messages.

The MPMA experimental setup is generally similar to the system used in the first PSaD multicasting demonstration. The ten PSEs are arranged in the proposed multistage topology, with two PaR stages followed by three PaM stages. The routing control logic in the 2×2 PaR stages is as described above, using a two-bit control input (Figure 4.20b): the frame bit and the address bit. When the address bit is low, the packet is routed to the upper port, and when the address bit is high, the

packet is routed to the lower port. In the case of the PaM stages (Figure 4.20b), the routing control logic was modified by reprogramming the CPLDs in the experiment to implement the packet multicasting routing. The programmable logic also uses a two-bit control input: when the address bit is low, the packet ingressing on the 2×2 PSE is transmitted across (*i.e.* upper input port to upper output port, and lower input port to lower output port), and when the address bit is high, the packet is transmitted to both of the output ports (*i.e.* upper/lower input port to both output ports).

To demonstrate the multicasting operation using MPMA, an experimental predetermined pattern of multiwavelength optical messages is generated and injected in the 4×4 switching fabric test-bed. In addition to the eight 10-Gb/s payload channels, each message utilizes a six-wavelength control header, comprising of one frame bit and five address bits: one address bit for each routing stage in the fabric. Depending on the high/low levels of the address bits encoded by the transmitter in the packet header and the type of logic encoded within the stage, the CPLD within the 2×2 PSE either routes or multicasts the wavelength-striped packet by gating on one or two SOAs. The payload information for the optical packets is generated similarly to the above demonstration, using eight CW DFB lasers. The wavelength channels are concurrently modulated at 10 Gb/s with a single $LiNbO_3$ modulator with a 2^{15}-1 PRBS NRZ-OOK signal. The eight-channel modulated payload is then transmitted to a discrete SOA that is gated using the ParBERT. The six control headers are generated separately and the ParBERT provides the corresponding addressing for the packet routing and multicasting stages. The control headers are then multiplexed with the payload channels, creating the 8×10-Gb/s optical packets using 192-ns timeslots and supporting 179.2-ns duration

messages.

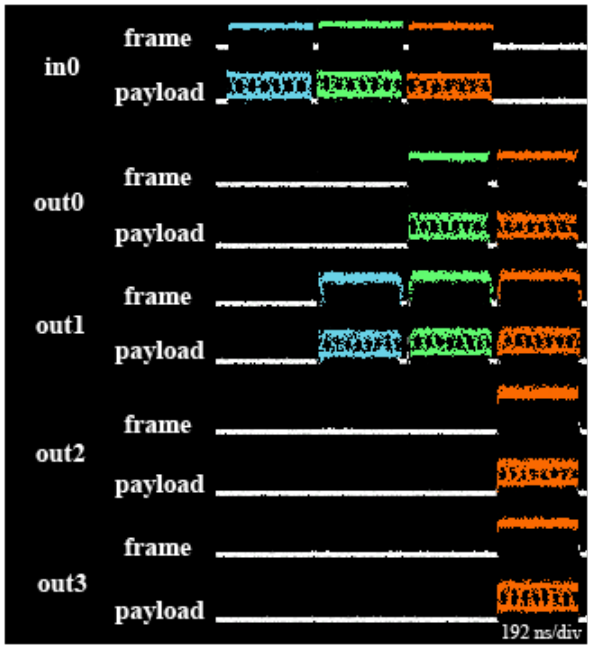

Figure 4.21: MPMA Waveforms - Optical waveform traces corresponding to the experimental optical packet sequence, which exemplify the multicasting operation executed by the realized topology.

The packets are injected in the switching fabric, and are distributedly routed (in the PaR stages) or multicasted (in the PaM stages) by the 2×2 PSE according to the encoded optical headers. All the unique header combinations are shown to demonstrate a one-way, two-way, or four-way multicast. The optical waveforms corresponding to ingressing and egressing optical packets are shown in Figure 4.21, providing the traces for the frame bit and one modulated payload channel. In the first active timeslot, a packet (colored blue in Figure 4.21) is injected in the fabric test-bed and according to its encoded addressing information, is transmitted to one output port. This first packet has address information 00000, routing the wavelength-striped message from

in0 to out1 according to the distinct PaR and PaM routing logic tables. This packet emerges from the output out1. In the second timeslot, the next packet (colored green in Figure 4.21) has an encoded address of 00001, which routes the message simultaneously from in0, to both out0 and out1. The multicasting is initiated by the fifth PaM stage, since the fifth address bit is high, indicating that the multiwavelength message should be transmitted to both output ports of the fifth stage of the fabric test-bed. It can be seen that the green packet is successfully multicasted and egresses from the desired destinations accordingly. In the third timeslot, a packet (colored red in Figure 4.21) with address 00011 is injected in the test-bed. The encoded control header shows that the packet wishes to be multicasted to all four available output ports. The last two (fourth and fifth) bits are high, indicating that the 2×2 PSEs in the fourth and fifth PaM stages will multicast the packet to both of their output ports. This allows an optical multicast (or broadcast) of the multiwavelength packet from in0 to all four output ports; accordingly, the red packet is effectively transmitted to all four outputs simultaneously.

The waveform traces show that all of the 8×10-Gb/s multiwavelength messages are confirmed correctly routed, egressing at the destinations as designated by the control address headers. The experimental demonstration verifies that the second MPMA topology seamlessly enables unicasting to one output, multicasting to two ports, in addition to broadcasting to all four outputs.

The packet analysis system is typical in setup, and similar to other experiments, using a DC-coupled 10-Gb/s *p-i-n* photodiode and BERT that is synchronized with the gating ParBERT. The BERT is gated over more than 80% of the packet duration (over

150 ns of the packet length). BER measurements confirm the error-free transmission of the supported wavelength-striped packets at the output of the fabric test-bed. BERs less than 10^{-12} are achieved on all eight of the payload wavelength channels, obtaining no errors in 10^{12} bits. Sensitivity curves of the eight payload wavelengths of a packet emerging from the five-stage test-bed are given in Figure 4.22. The BER curves show a power penalty of 2.5 dB for the five-stage fabric (taken at a 10^{-9} BER), resulting in an approximate 0.5-dB penalty for each SOA traversal; this corresponds to a similar penalty performance as in the first packet multicasting demonstration and is within experimental error. The insets show the optical eye diagrams for a single 10-Gb/s channel (at λ=1556.6 nm) at the fabric input and output.

Thus, this demonstration clearly shows that the proposed MPMA topology can successfully realize the multicasting functionality with error-free operation, with potentially improved scalability and reduced cost as compared to the initial PSaD design. It can be noted that this multistage switching design gives rise to greater probability for packet contention. In the future, a control plane may be required to achieve non-blocking lightpaths for the optical messages. However, the management and scheduling of the packet multicast from the higher network layers is considered difficult (NP-hard) [122]. This design aims to reduce the control complexity on the optical layer and does not address the difficult multicast scheduling problem.

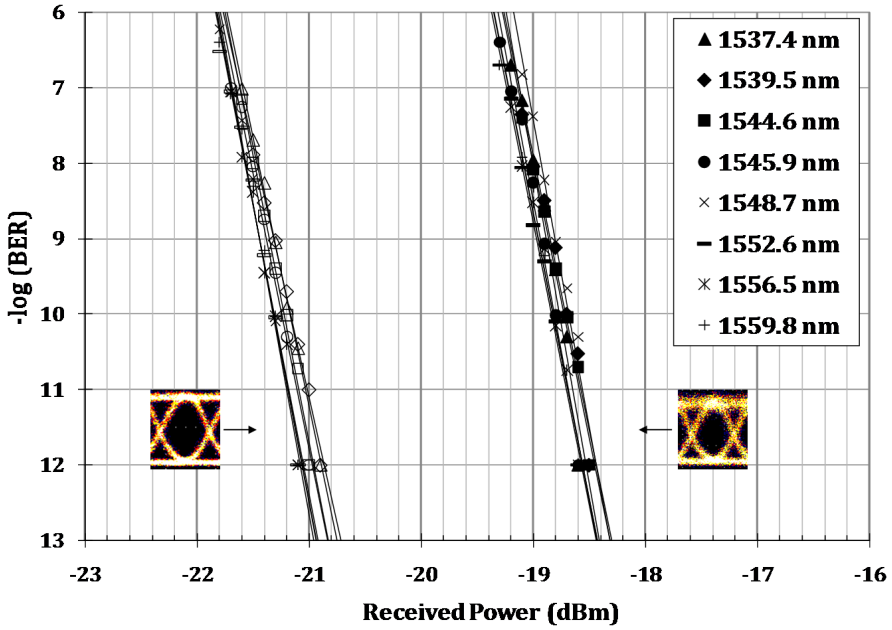

Figure 4.22: MPMA Sensitivity Curves - BER sensitivity curves corresponding to the multicasting operation, taken for all eight payload channels: open data points refer to measurements for packets at the input of the fabric test-bed, while filled data points correspond to measurements taken for packets at the output of the test-bed. 10-Gb/s eye diagrams for the input and output packets are provided as insets (λ=1556.6 nm).

4.7 Analysis: Comparison of Multicast-Capable Designs

Finally, it is worthy to perform a side-by-side comparison of the PSaD and MPMA designs in terms of the hardware required to implement each of these designs in this test-bed environment. The following comparative analysis assumes the implementation of a 4×4 (*i.e. N=4*) optical packet switching fabric using the current 2×2 PSE design, wherein each PSE extracts two control header bits (one frame and one address bit). The following analysis also uses the minimum number of components (*i.e.* to explore the hardware cost) required by both the PSaD and MPMA approaches in the exemplary case of enabling a packet multicast fan-out of four. For instance, the PSaD experimental demonstration here only interconnects *N=4* ports with a maximum of *M=2*-way multicast; the first parallel optical switch consists of a three-stage switch topology and the second is a two-stage switch topology. In this comparison case study, a PSaD architecture is assumed that connects *N=4* ports to enable a possible *M=4*-way multicast, with each of the parallel internal optical switches consisting of two-stage topologies (Figure 4.23a). In this way, the PSaD design is assumed to use the minimal number of components with a reduced additional hardware cost to support the multicasting application. The PSaD architecture also requires several $1\times M$ and $M\times 1$ SOA-based switches at the fabric input and output, respectively, to compensate for the splitting and combining insertion losses. For the assumptions associated with the MPMA design, the experimentally-demonstrated architecture is used, with five stages of 2×2 PSEs (Figure 4.23b). The five-stage implementation is required to provide both the unicast and multicast capabilities for a 4×4 fabric.

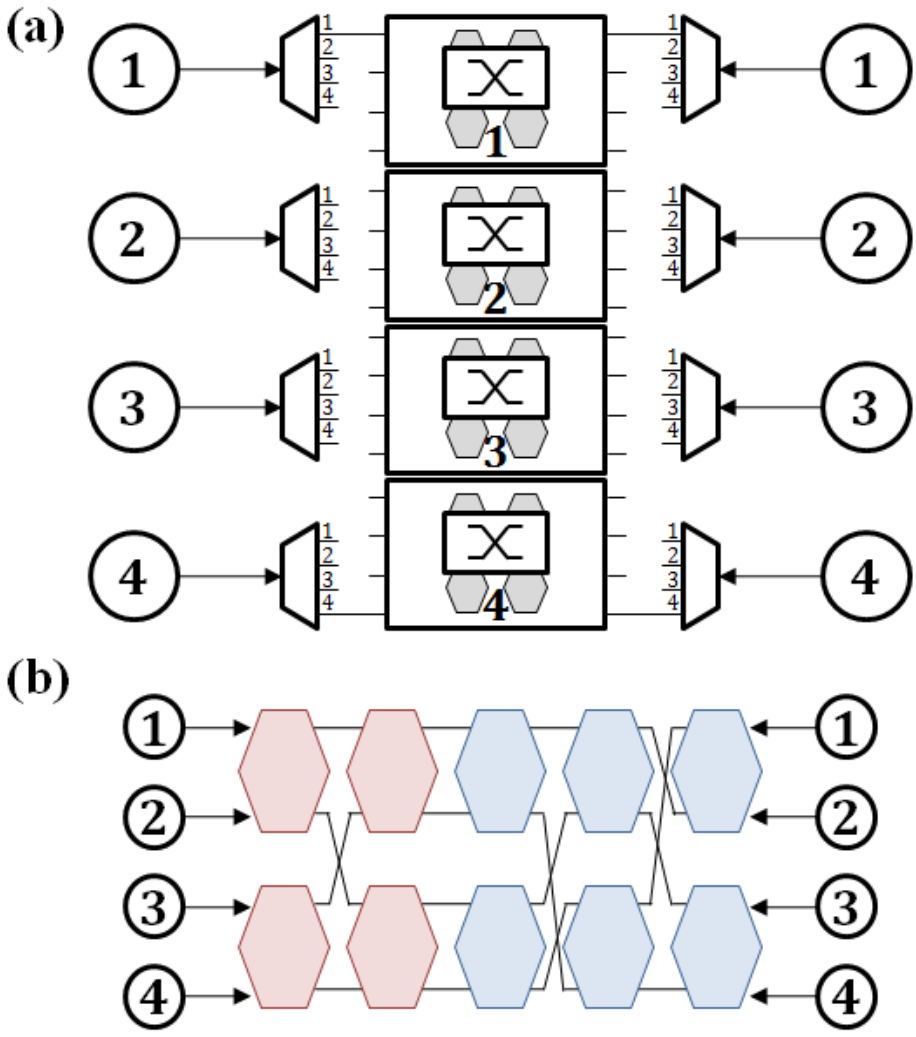

Figure 4.23: Multicast-Capable Design Comparison Case Study - Diagrams of the (a) PSaD and (b) MPMA topologies assumed in the exemplary comparative analysis.

Table 4.1: Assumptions and results of the specific case study to compare the two discussed multicast-capable designs.

Parameter	PSaD	MPMA
Input/output ports (N)	4	4
Maximum multicast fan-out	4	4
Number of parallel switches (M)	4	N.A.
Number of 1:4 switches	8	N.A.
SOAs required for implementing switches	32	N.A.
Total number of routing stages in fabric	8	5
Total number of 2×2 PSEs	16	10
SOAs required for PSEs	64	40
Total number of SOAs	96	40

Table 4.1 depicts the results of the comparative analysis, showing the hardware cost of the specific case of realizing a 4×4 optical switching fabric that can support a four-way multicast. The primary hardware metric is the number of SOAs required by the multicast-capable switching fabric using both the PSaD and MPMA approaches. Since the M-way PSaD design requires more additional PSEs, as well as 1:4 SOA-based switches, the number of SOAs needed to implement the PSaD design (*i.e.* 96) far outnumbers that of the optimized MPMA design (*i.e.* 40). It can thus be seen that for this representative example, MPMA indeed exhibits a reduced cost in hardware to support the equivalent multicasting fan-out. If the future optimized 4×4 PSE is used instead of the current 2×2 PSE design (*i.e.* a PSE version that is capable of extracting a one-bit frame and a two-bit address per PSE input port), the MPMA design implementation would exhibit further reduced hardware costs with much simpler routing control logic.

4.8 Closing Remarks

In order to effectively meet the high-bandwidth demands of the future Internet with ultralow cost, it is evident that greater functionality, intelligence, and capabilities will be necessary on the physical layer. Providing the optical layer with greater switching functionality based on higher-layer networking approaches will yield more effective network routing with minimal O/E/O conversions. Further, lower energy consumptions can be achieved by migrating these functionalities to the optical layer. This research effort aims to produce these intelligent, "adaptable" optical pipes to support larger bandwidths with the added ability to dynamically manipulate optical switching on a packet-granular scale [29].

Chapter 5

Dynamic Cross-Layer Platform

In this chapter, the bulk of the author's contribution in developing a cross-layer communications platform for next-generation optical networks is described. This includes the development of a message control interface, advanced packet protection switching capabilities, explorations of QoS-based multicasting, as well as extensive work on performance monitoring that can provide feedback signals to a cross-layer control plane.

Cross-layer communications facilitates the dynamic and intelligent management of packet transmission based on the packets' physical-layer quality (*i.e.* QoT) and their higher-layer priority attributes (*i.e.* QoS). The vision is to be able to efficiently control the optical layer at packet- or flow-level granularities, in order to enable more flexible, intelligent resource allocation through global cross-layer optimization algorithms. Cross-layer networking aims to create the more intelligent management of optical data by providing feedback from the physical layer to the higher-level routing and application layers. In this way, the cross-layer routing algorithms can support a

more dynamic physical layer, based on the constraints imposed from the higher network layers. This bidirectional signaling platform will enable an Internet infrastructure that can adapt to both users' requirements as well as the performance of high-data-rate optical signals [29]. The new cross-layer network designs and architectures will be essential to addressing these exploding transmission bandwidths by providing flexible networking and energy-aware capabilities [30].

This concept of cross-layering lends itself directly from the wireless domain. Cross-layer communications has been used extensively in wireless networks to optimize network scheduling and resource allocation with respect to power utilization [131, 132, 133, 134]. Initial notions of cross-layer networking have been studied to optimize optical network performance (as mentioned above in Chapter 2), in order to enable advanced traffic engineering and impairment-aware routing algorithms with an optical control plane (OCP) [135].

This work centers specifically on the design of a bidirectional cross-layer exchange infrastructure that can utilize dynamic real-time OPM measurements via extraction of the packets' and/or flows' BER or other signal quality metrics. As data streams propagate on the optical layer, the physical signals may be affected by degradations in OSNR, crosstalk, dispersion (*i.e.* CD or PMD), or optical nonlinearity, all contributing to a reduced BER. Various introspective technologies can detect signal degradations in real-time and feedback performance information to higher layers to help ensure network reliability and robustness, which is particularly important as optical data rates continue to scale to meet the high-bandwidth demands.

The envisioned future infrastructure will support flexible packet routing and

protection capabilities at these required higher data rates. By monitoring the quality of multiple packets that constitute a single optical flow at a packet granularity, the performance of the flow can be ascertained and flow-level rerouting can be actuated. Thus, this thesis proposes that flow control on the optical layer can be triggered by performance monitoring that is executed on a packet-by-packet basis. Our work focuses on providing more dynamic network management that can efficiently account for optical-layer performance, and control the optical packets that make up larger optical flows. By realizing this programmable control of the optical layer, dynamic access to the physical layer can be leveraged to create cross-layer impairment-aware network control schemes, more efficient network resource allocation, and cross-layer optimization algorithms.

This work envisions an architectural design as shown in Figure 2.6, which can leverage optical devices as accessible components, while dynamically optimizing performance and power consumption in a cross-layer optimized way [35, 115]. This will provide a high level of control over multiwavelength optical messages, ideally with a packet-level granularity. The proposed cross-layer mechanisms must also be kept very simple to allow for fast processing (ideally packet level) to comply with the fast packet switching requirements. The platform allows for the programmable optical layer to dynamically interact with higher network layers, creating a bidirectional cross-layer information flow. In this work, the optical physical layer is no longer assumed to be a static black box, but rather a programmable switching fabric that is network- and application-aware, and provides a monitoring functionality to the higher network layers. The switching fabric comprises of a hybrid opto-electronic fabric to achieve fast packet-

scale all-optical switching. This functionality is not currently available with standard commercial optical system technology, and further research is required to motivate network architectures and optical substrates that can lead to viable integrated cross-layer solutions.

As discussed in Chapter 2, the cross-layer signaling platform that is presented here allows the optical switching fabric to interact in a highly dynamic fashion with the higher network layers. IP-layer QoS requirements can flow down to the physical layer to affect optical-layer handling, while QoT (PM or OPM) metrics can be extracted and these measurements flow upward for higher-layer influence. This chapter explores the development of the required cross-layer designs and routing control algorithms based on these integrated optical devices, real-time physical-layer measurements, and incorporation of varying QoS protocols. These routing schemes can dynamically optimize physical-layer switching, enabling a deeper exposure of the physical-layer substrate. Additionally, the ultimate platform will endeavor to incorporate, drive, and exploit the emerging revolutionary and heterogeneous advances in physical-layer technologies, by allowing the integration of novel, flexible optical devices and monitoring subsystems directly in the physical layer to provide substantial performance gains for the overall network: this includes both OPM modules and switching technologies. The network will be able to holistically and intelligently recover from failures, manipulate optical data based on the signals' performance, and efficiently allot bandwidth.

5.1 Message Control Interface

As a first stab at the notion of cross-layering, this section presents an initial signaling platform consisting of a message injection control and queue management technique for optical packet switching fabrics that accepts input from the switching fabric itself [136]. If messages are dropped within the switching fabric, a short cross-layer control signal is sent to a deployed message injection control interface for packet retransmission. This demonstration of cross-layer communications involves the interoperation of an OPS fabric with an optical input/interface buffer. In a time-slotted manner, cross-layer signaling is employed between the input buffer and fabric to dynamically reroute dropped payload packets with multiple wavelength-striped 10-Gb/s channels.

Although OPS can offer dynamic management of the vast growth of high-throughput traffic in next-generation routers [24], a significant challenge to implementing OPS fabrics is resolving contentions, which may occur at a given PSE within the fabric as multiple packets attempt to utilize the same output link. In a complementary electronic network, contentions are straightforwardly addressed by buffering packets and forwarding them once the contending path is freed. However, owing to the absence of practical optical buffering elements, simple contention resolution is difficult to achieve in OPS fabrics. This shortcoming can be partially mitigated by employing small-capacity optical packet buffers at the input. These buffers can accept back pressure due to contentions and control the traffic injected into the fabric [137]. For instance (as demonstrated here), the optical buffers can function as queues that store a copy of the input packets preceding injection into the optical fabric. Once a packet acquires a contention-free path and is successfully

transmitted, it is then discarded from the queue. In the case of unsuccessful message transmission (*i.e.* packing dropping within the fabric), the input buffer can reattempt fabric injection with the copy of the message stored in the queue.

An optical packet buffer architecture has been previously presented [138, 139]. Furthermore, the architecture's flexibility and reconfigurability with respect to fabric congestion control has been shown via the demonstration of active queue management [140, 141]. Here, the basic buffer architecture has been adapted to realize the functionality of packet injection control at the interface of an implemented OPS fabric. The OPS test-bed utilizes the basic fabric architecture outlined in Chapter 3 (also in [25, 26]), which consists of a switching fabric comprising of 2×2 PSEs. The fabric does not employ optical buffering within its PSEs, thus messages are dropped upon contention. The ack protocol allows for a drop-detection mechanism in which a short optical ack pulse is sent from the receiving port upon successful transmission. Retransmission can then occur with minimal latency and reduced penalty due to dropping messages.

Here, the interoperability between the implemented fabric interface packet buffer and a 4×4 OPS fabric test-bed is experimentally demonstrated [142]. The interface buffer actively queues packets prior to injection into the network (Figure 5.1). The ack pulses are further leveraged to provide a means of cross-layer signaling at the buffer-fabric interface, thus mitigating unsuccessful transmissions through the test-bed. The buffer discards its copies of correctly routed packets, while dynamically retransmitting packets dropped due to contention within the fabric. High-bandwidth optical packets containing 6×10-Gb/s wavelength-striped payloads are transparently processed at the

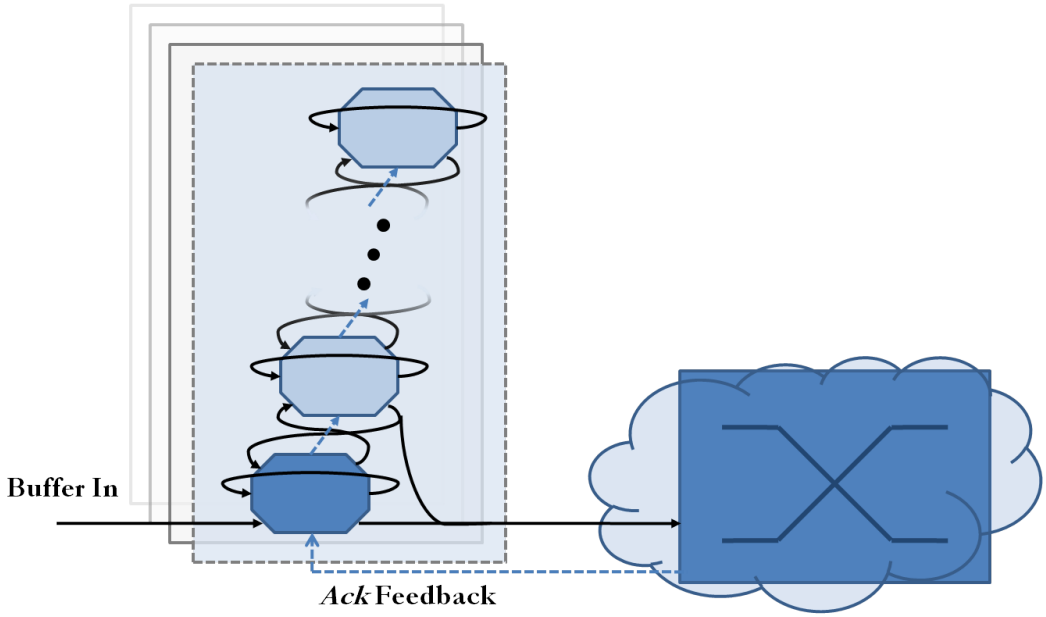

Figure 5.1: Control Interface Architecture - Cross-layer architecture for signaling communication interface between optical input buffer and OPS fabric.

interface buffer and correctly routed through the OPS test-bed.

5.1.1 Optical Buffer Architecture

The basic optical packet buffer architecture is composed of identical building-block modules organized in a cascaded hierarchical structure (Figure 5.2). Each module uses a 3×3 SOA-based switch and can buffer a single packet of fixed length in a FDL. Using a time-slotted approach, packets of fixed lengths enter the buffer via the root module. If the buffer is empty, the packets are stored in the root's FDL; otherwise, packets are propagated upward along the cascade until an unoccupied module is encountered. In a typical implementation, first-in first-out (FIFO) ordering ensures that the age of a given packet corresponds directly with its position in the cascade (Figure 5.2). Additionally,

reading packets from the buffer is independent of the write process. An electronic read signal is transmitted to the root module, which subsequently forwards the contents of its FDL to the output. The read signal is then regenerated and retransmitted through the cascade, advancing the stored packets incrementally towards the root. Each module is a self-sufficient unit requiring no central management. Increasing the total buffer capacity is realized by connecting additional identical modules to the end of the cascade.

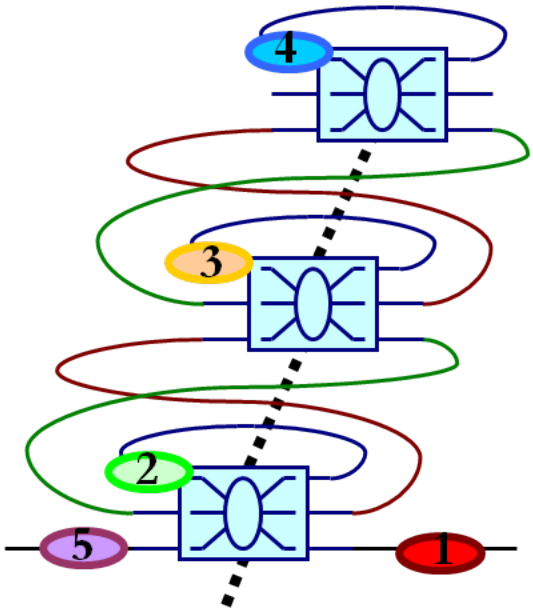

Figure 5.2: Optical Packet Buffer Architecture - Example buffer architecture with 3×3 SOA switches. FIFO functionality is shown, with the numbers corresponding to the relative age of the packets (1 is oldest, 5 is newest); dashed lines depict the electronic control read/write signals.

In order to implement the switching fabric interface buffer, the basic architecture was modified via the programmable modules to offer interoperability with the OPS network (Figure 5.3). The adapted buffer stores and transmits the oldest

unacknowledged packet at each timeslot until an optical acknowledgement pulse is sent by the network. Once a packet is successfully transmitted to the output port, an ack is sent to the buffer. The Ack Translator produces an electronic ack_in, which is received within the same timeslot, replacing the read signals in the prototypical design. Upon reception of an ack, the currently transmitting packet is discarded from the buffer's queue in the following timeslot and the next packet in the buffer is injected into the network. If a packet is dropped due to contention, the buffer will dynamically retransmit the packet until it is successfully transmitted to the output. In this way, the physical-layer ack pulse is leveraged as means of feedback cross-layer signaling between the interface buffer and the fabric. To provide immediate egression for serially acknowledged packets, the buffer architecture provides an additional output packet pathway from the first module.

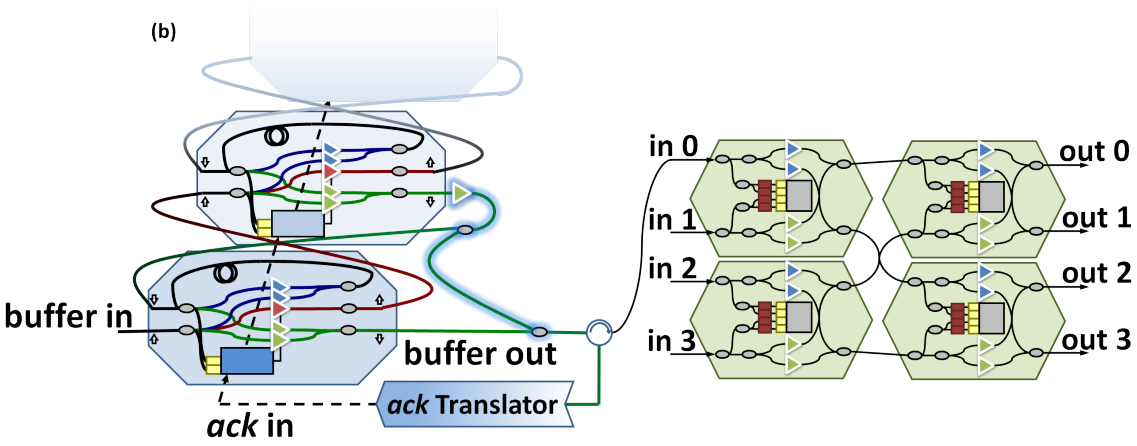

Figure 5.3: Message Interface Setup - Experimental setup corresponding to a two-module buffer with OPS fabric test-bed.

5.1.2 Experimental Demonstration

To demonstrate the interoperability of the network interface buffer with the OPS test-bed, optical packets are first injected into the buffer and are subsequently stored in the buffer's queue until a successful network transmission occurs. The system's broadband transparency is illustrated through a 6×10-Gb/s wavelength-striped packet format. The payload is segmented and modulated at 10 Gb/s on six additional wavelengths across the ITU C-band.

A two-module experimental prototype of the interface buffer (Figure 5.3) is built for these demonstration experiments. The decision logic is synthesized in a high-speed Xilinx CPLD, with two distinct versions of the logic truth table: one pertaining to the root module and another for subsequent, higher-order modules. This allows for the necessarily asymmetric behavior of the root module with respect to the higher-order modules. Each building-block module is comprised of commercially available components: the aforementioned CPLD, five SOAs operating as switching elements, and two 155-Mb/s *p*-*i*-*n* photodetectors. No optical filters are necessary in this implementation. The 4×4 experimental network test-bed (Figure 5.3) comprises of four 2×2 PSEs, also realized with discrete components. As described in Chapter 3, the network's PSEs decode control information by filtering the two header bits (frame and address); using the control information, the CPLD then gates the node's four SOAs to appropriately route the packets.

Both the buffer and fabric systems leverage SOAs as their switching elements, allowing for the compensation of insertion losses that arise from the passive coupling elements internal to each node and buffer module. In this way, each SOA hop

contributes no net gain or loss, and packet longevity is maintained. For the interface buffer implementation, an additional SOA is required at the output of the second module; this SOA acts as a simple amplifier for the purpose of gain equalization between the root and first buffer modules. The implemented system supports 128-ns timeslots, containing 115.2-ns duration packets with 10-Gb/s modulated data on six payload wavelengths. The packets are modulated by a single modulator with a 2^7-1 PRBS at 10 Gb/s in NRZ-OOK format. Due to the transparency of the constructed components, the modulation format of the payload data is not a limiting factor. Cross-layer signaling ack pulses are generated by a ParBERT to inject packets from the buffer.

5.1.3 Results

Figure 5.4a depicts the experimental optical packet sequence as the packets are injected in the buffer-fabric system. Figure 5.4b shows the input and output optical waveforms for packets emerging from the integrated buffer-fabric operation. All the packets from the fabric are shown to be correctly routed. Packet A is first stored in the buffer and is simultaneously injected into the network; it is successfully transmitted and thus discarded from the buffer. Two timeslots later, B is similarly injected into the buffer and fabric. Simultaneously, another network port injects D, causing contention between B and D. In this case, D is received at the network output, while packet B is dropped. Consequently, the buffer does not receive an ack and then re-injects its stored copy of B. Packet C also appears at the buffer input and thus is stored in the buffer. The second transmission of B is successful, and in accordance with FIFO ordering, C is injected into the network in the following timeslot and is successfully transmitted to

the network output.

Figure 5.5 portrays the input and output eye diagrams for one 10-Gb/s payload wavelength (λ = 1558.31 nm) for C, which undergoes six SOA hops: the greatest number of SOAs experienced by any packet. The packet is amplified at the network output using an EDFA, filtered with at tunable grating filter, sent to a VOA, and then received by a *p-i-n* photodiode with TIA and LA. The received electronic signals are then transmitted to the BERT that is synchronized with the packet gating signal. The modulating bit pattern is driven by the PPG. BER measurements show that the packets are received from the output error-free with BERs less than 10^{-12} on all six payload wavelengths. To further support the system's enhanced operation, Figure 5.6 presents BER curves at 10 Gb/s for packet C (traversing six SOA hops). No error floor is observed and the power penalty is evaluated at a BER of 10^{-9} as 3.5 dB, indicating a power penalty of 0.6 dB per SOA hop.

This section presents the initial investigations into the feasibility and functionality of cross-layer communications as demonstrated via the joint operation of an injection control buffer with an OPS fabric test-bed. Multiwavelength optical packets are transparently processed by the interface buffer and dynamically rerouted through the test-bed. Wavelength-striped optical packets with 6×10-Gb/s payloads are routed, with error-free transmission on all payload wavelengths (BERs less than 10^{-12}). This exploration exemplifies the potential for enhanced network performance through the dynamic interoperability between an optical packet injection control buffer and an OPS fabric.

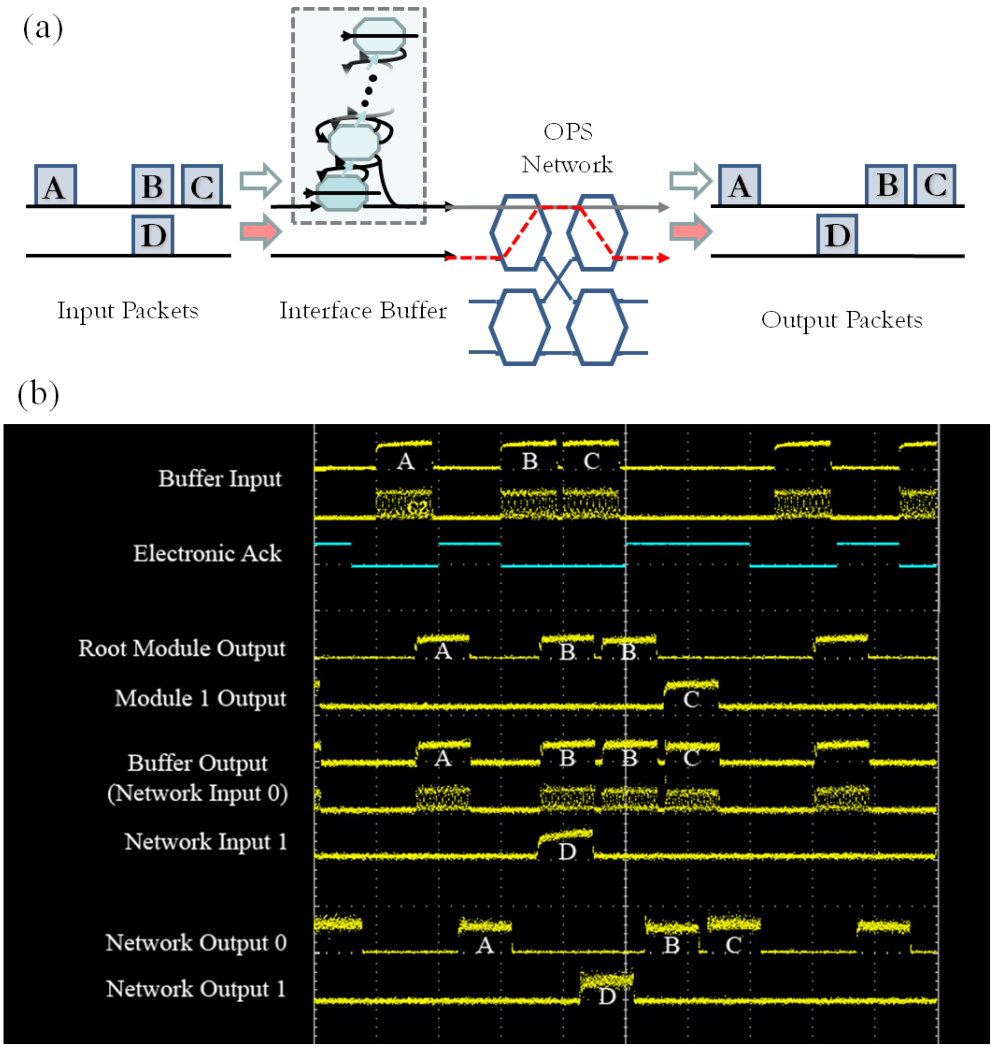

Figure 5.4: Message Interface Waveforms - (a) Diagram depicting the experimental packet sequence with the buffer and fabric. Contention occurs between packets B and D; thus, B is retransmitted at a later timeslot; (b) Optical waveform traces for the buffer and network input and output signals.

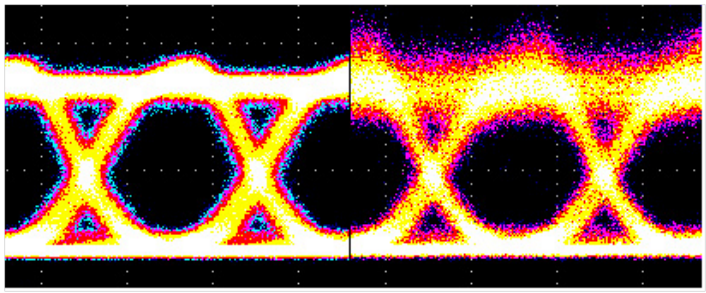

Figure 5.5: Message Interface Eye Diagrams - 10-Gb/s input (left) and output (right) optical eye diagrams of packet C, which undergoes six SOA hops.

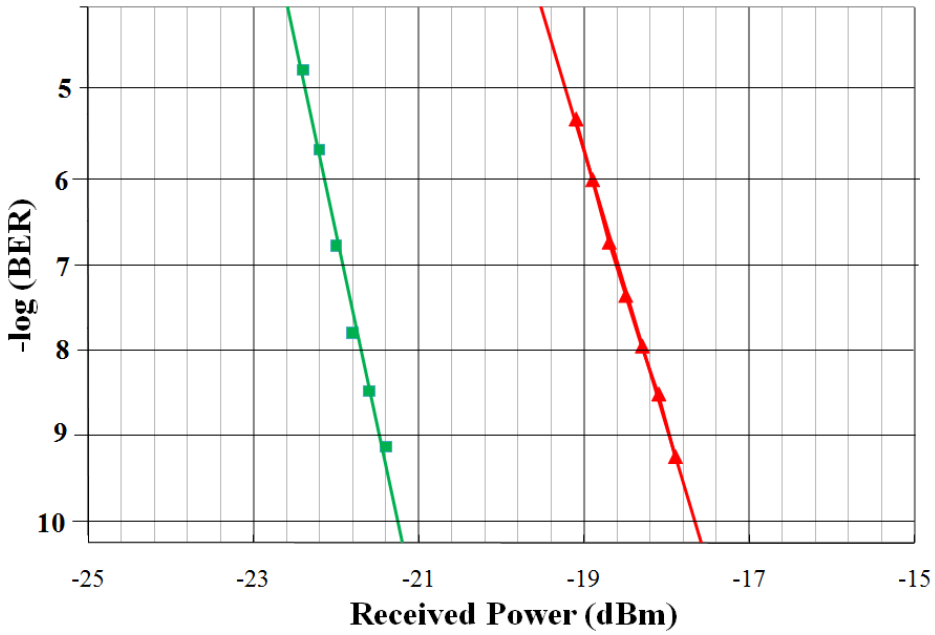

Figure 5.6: Message Interface Sensitivity Curves - 10-Gb/s BER sensitivity curves for packet C, experiencing 6 SOA hops: red data points correspond to output through measurements, while green data points correspond to input back-to-back measurements ($\lambda = 1558.31$ nm).

5.2 Packet Protection Techniques

The following section of this thesis explores an advanced packet protection technique that can be used to showcase the benefits of cross-layer communications. The proactive packet protection (PPT) switching mechanism is experimentally demonstrated using a cross-layer platform on the OPS fabric test-bed, with accompanying packet network simulations. The cross-layer design will drive the development of optical systems technology in parallel with the network architecture, where packet-level OPM devices can be directly embedded in the physical layer. The bidirectional flow can support differentiated QoS requirements to flow downwards into the physical layer, such that the data's QoS class can be invoked directly on the optical layer. Correspondingly, the OPM devices can extract the real-time physical-layer performance, possibly indicating isolated signal impairments, and send these measurements upward in the network stack. This cross-layer platform may grant future networks the ability to improve overall network management, operation, and performance based on the packet's QoS and optical signal QoT.

As shown in Figure 5.7, this work focuses on a cross-layer network architecture that uses physical-layer PM devices, for example via the extraction of the packets' BER. By monitoring the flow at a packet granularity (*i.e.* monitoring the quality of packets constituting a flow), the performance of the data stream can be ascertained, resulting in the actuation of flow-level optical rerouting. The PPT scheme, proposed by Gerstel *et al.* in [143], is investigated that allows the control of data traffic flows through physical-layer PM in combination with higher-layer QoS parameters. PPT allows for a degrading BER to be detected by cross-layer control logic. As shown in

Figure 5.8, if the packets of a flow cross a predefined BER threshold (BER_T), flow rerouting is triggered to proactively avoid packet loss of high-QoS flows in case the BER further degrades beyond the correction threshold BER_E of the underlying forward error correction (FEC) mechanism [144]. Different QoS classes may use different, flow-dependent BER thresholds for flow rerouting.

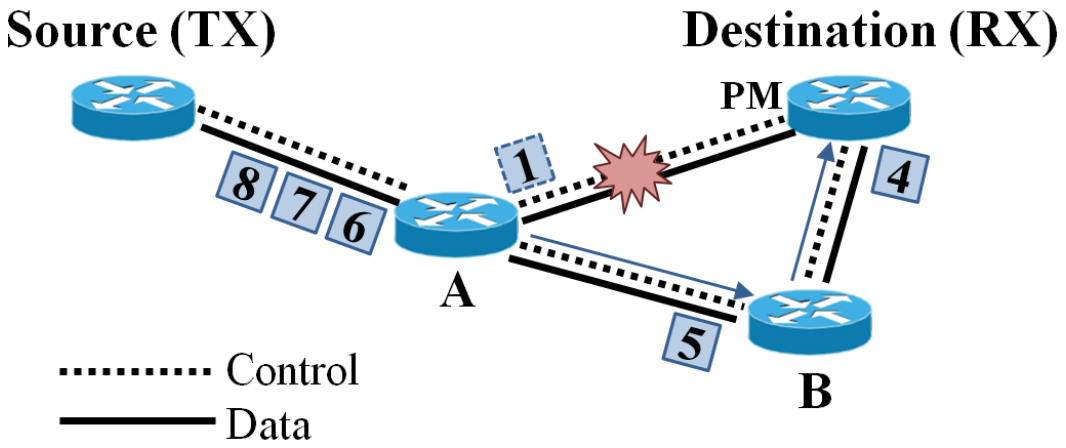

Figure 5.7: PPT Network Architecture - Complete envisioned architecture depicting the rerouting scheme for a flow of data packets.

Figure 5.7 shows a representative network architecture where the routing of flows takes into account the physical-layer performance. A transmitter (TX) is sending packets to a receiver (RX) via an intermediate router A. The PM device at the RX sends a control signal to router A if it deems the link connecting A and RX to be impaired. This triggers rerouting of high-QoS packets via router B to RX well before the link BER becomes too high to be handled by the underlying FEC. The cross-layer rerouting decision can also be based on a higher-layer compromise between the expected packet loss probability and the enhanced delay and delay jitter (*i.e.* QoS) due to the longer reroute path via B. Thus, on a flow-by-flow basis, the scheme allows for the

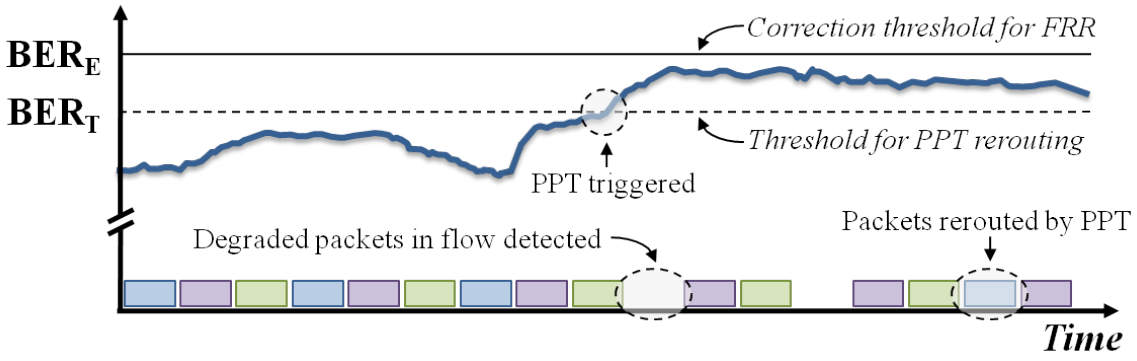

Figure 5.8: PPT Diagram - Diagram of PPT switching, showing packets constituting a flow with respect to BER variations. As the BER increases above BER_T, the PPT scheme is triggered and packet loss if avoided.

switching and protection of packets based on the optical signal quality in combination with higher-layer QoS parameters. In the context of this work, the chosen PM metric is a BER measurement; however, the implementation is not limited to this quantity. The integration of novel optical technologies that can achieve high data rates may also allow flow- and packet-level physical-layer measurements [63, 64].

The PPT exploration is two-fold, including both an experimental component and a simulation investigation. This work first presents an experimental demonstration of a programmable-logic-controlled optical packet switch fabric test-bed, demonstrating some key technologies needed to implement a fast hybrid optical/electronic switch with cross-layer PPT mechanisms [115]. The OPS test-bed supports 8×10-Gb/s wavelength-striped optical packets, *i.e.* messages with 10-Gb/s data payloads that span 8 wavelengths using WDM, with additional wavelength-striped control headers. All wavelengths are routed together cohesively from the network input to its output. The packets are switched and rerouted based on input from cross-layer control logic,

(involving both signal quality parameters and packet QoS information). Transmission through the test-bed is obtained with measured BERs of less than 10^{-12}, and the performance of the system is quantitatively analyzed.

Using ns-2 simulations [145] and newly developed ns-2 modules (discussed further in Fidler *et al.* [146]), the proposed cross-layer PPT technique is then compared with a layer-2 fast-reroute (FRR) mechanism [147]. FRR corresponds to layer-2 switching that is in response to hard failures on nodes and links. In the case of FRR, as shown in Figure 5.8, the rerouting of a data flow to a protection path is triggered when the received BER exceeds BER_E, the correction threshold of the supported FEC in the receiver (*e.g.* when the BER $\geq 2\times10^{-3}$ [144]). With PPT, the rerouting process is triggered at a lower predefined BER threshold $BER_T < BER_E$ (here, $BER_T = 10^{-4}$), promising near-hitless protection (*i.e.* no packet loss) for relatively slow impairment dynamics [143]. Since today's discrete-event packet network simulation environments do not support realistic physical-layer performance variation models, new simulation modules [145, 146] are created to incorporate physical-layer BER variations with varying impairment timescales. The simulations aim to analyze dropped packets by allowing incorporation of dynamic BER variations. These results show that PPT yields throughput gains and decreased packet-loss rates, depending on impairment dynamics and network size.

5.2.1 Experimental Demonstration and Setup

To enable the proposed PPT mechanisms in future networks using hybrid opto-electronic packet routers, this work demonstrates some key technologies using the

router architecture shown in Figure 5.9. The system includes a 1+1 protected optical packet switch fabric, cross-layer control logic, PM devices, and subsystems to extract the flows' QoS [115]. The switch fabric was originally designed as multicast-capable [128]. As shown in Figure 5.10, the 1+1 protected 4×4 optical switch fabric (OPS1 and OPS2) uses the 2×2 non-blocking PSEs as its basic building blocks. The fabric is built from discrete, commercial off-the-shelf components, such as SOAs, low-speed *p-i-n* photodetectors, passive optical components and filters, and high-speed digital electronic circuitry. The test-bed is constructed using ten such 2×2 PSEs. A three-stage and a two-stage 4×4 network switch are implemented as multistage banyan topologies, as indicated in Figure 5.10.

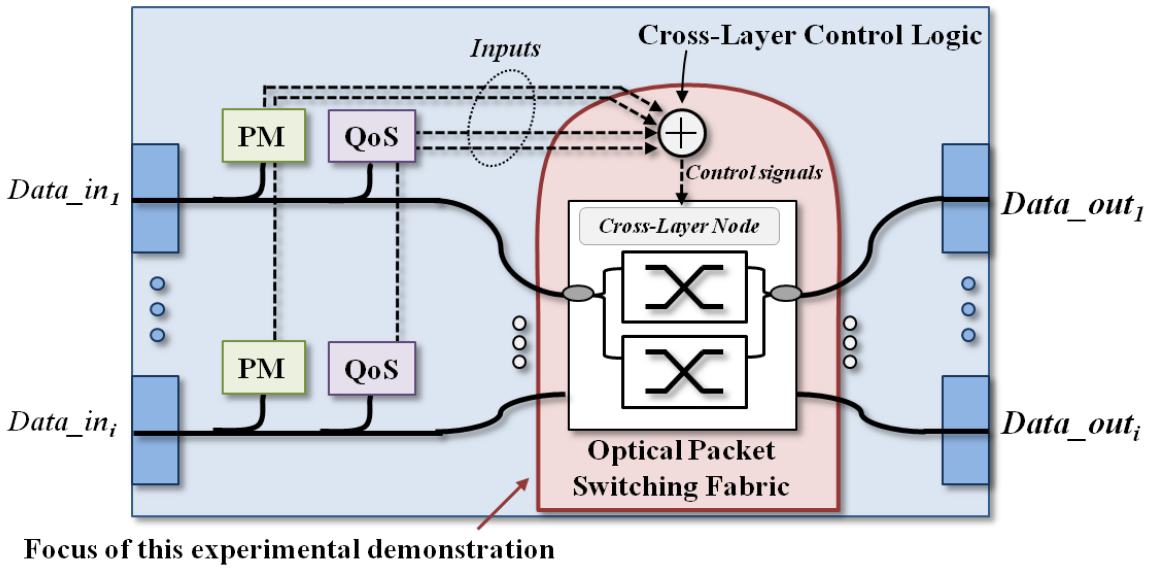

Figure 5.9: PPT Cross-Layer Network Node - Block diagram of a possible network node architecture: the red region corresponds to the focus of this experimental demonstration.

In the experimental demonstrations highlighted in this section, all packets are

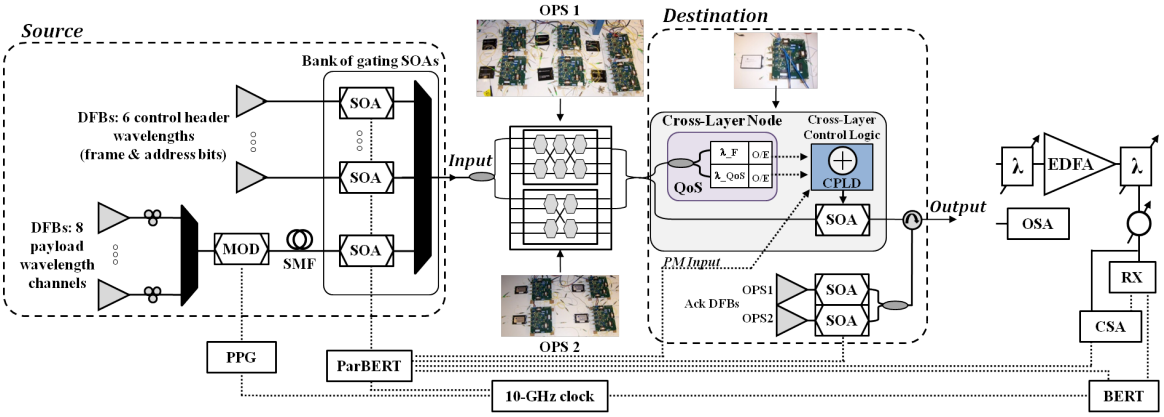

Figure 5.10: PPT Experimental Setup - Diagram of the experimental setup and the realized cross-layer enabled receiver design, with photographs of the optical fabric test-bed and cross-layer receiver. A detailed implementation of the receiver is given, showing how the routing decision can incorporate the packet's QoS in addition to measurement data feedback from a PM device.

synchronous and of equal length (*i.e.* "cells"). The system supports 128-ns timeslots and 115.2-ns duration packets. The 12.8-ns guard intervals allow for better time separation between packets in addition to mitigating the impact of finite rise and fall times of the SOAs. The timeslot duration corresponds to approximately 26 m of fiber, which is incorporated into the setup to achieve the necessary delays to accommodate electronic header processing without the need for extra payload buffering. Packets are routed all-optically through the switch fabric without prior request from the source nodes.

The wavelength-striped optical packet architecture is used in these experiments (see Chapter 3). The packet's control header information is encoded on a subset of wavelengths at a single bit per wavelength and is constant over the entire packet. The control header includes a frame signal {F}, denoting the presence of a packet and

spanning the entire length of the packet; the four-bit address signals (represented by {Ai, Aj}) denoting the packet's destination used for routing; and a QoS requirement bit, denoting the packet's priority class (as in Chapter 4). The packet's payload information is modulated at 10 Gb/s per data wavelength channel. With eight data channels, this results in wavelength-striped messages with 8×10-Gb/s payloads, and yields an aggregate message transmission bandwidth of 74 Gb/s, assuming a 7% overhead for FEC.

The CPLD makes a routing decision and gates the appropriate SOA gates, and the optical messages are then routed to their desired destination (or dropped upon contention). The routing logic is distributed among the individual switching nodes in OPS1 and OPS2 and is realized using multiple CPLDs located within the fabric's PSEs in the router. The ack mechanism is implemented to allow the short acks to be sent to the source to indicate successful transmission. Sources that do not receive acks can retransmit synchronously at the next timeslot. In this experimental implementation of the protection mechanism, contending packets are dropped. According to the QoS class, high-BER data packets are intentionally discarded upon reception by the receiver, with the aim of suppressing the ack and triggering rerouting with the packet protection scheme.

A pattern of optical packets is injected in the test-bed (Figure 5.10). The payload channels are generated using eight CW-DFBs ranging from 1539.6 nm to 1560.2 nm, with a minimum wavelength spacing between two adjacent channels of 0.8 nm. All eight lasers are combined using a passive optical coupler and then modulated simultaneously using a single $LiNbO_3$ modulator driven by a 10-Gb/s PPG. The signal carries a 2^{15}-1

PRBS in a NRZ-OOK format. The multiplexed payload channels then pass through a 24-km span of SMF to decorrelate the data. The control header signals are created using six separate CW-DFB laser sources at the appropriate wavelengths for the frame and address bits, including one frame signal at 1555.75 nm and five address headers, ranging from 1531.12 nm to 1550.92 nm. The control header and multiwavelength payload data channels are then gated using external SOAs and combined together, creating optical packets with the corresponding address encoding for message routing through the test-bed. The ParBERT is synchronized with the PPG and acts as a fast electronic signal generator to control the fabric addressing and packet gating using the external SOAs. The ParBERT is pre-programmed with sequences of optical packets for the experiment. The optical ack pulses are realized using a DFB laser at 1541.1 nm and are generated using the ParBERT gating on an external SOA. The packets' payloads are not synchronized with the PRBS pattern from the PPG. The packets are injected in the active input ports of the test-bed with two different QoS classes: high and low priority.

The PPT switching is exemplified for one data source/sink combination. The experimental system uses cross-layer control logic in a module shown as "Cross-Layer Node" in Figure 5.10, which is implemented at the receiving end of the test-bed with an additional SOA-based switch. It monitors the packets within an optical flow as they egress from the test-bed. The cross-layer control logic (depicted in Figure 5.9) is programmed as routing logic within the cross-layer node's separate CPLD, allowing flows to be proactively rerouted if the signal is degraded, or forwarded to the output port otherwise. The CPLD's switching scheme is also QoS-aware and thus is based

on both the packet's priority and BER parameters. Depending on the QoS and PM input, the SOA is gated on to either forward or discard high-priority messages. Flows which require a high QoS but show a high-BER are rerouted on a protection path, while low-QoS (regardless of BER) and high-QoS, low-BER messages are forwarded. A low BER indicates a signal quality below the predefined BER_T threshold for PPT, while a high BER denotes a quality above BER_T.

The system and logic allow for the BER measurement to occur over several consecutive packets that make up longer optical flows, as required by the frame format of the underlying FEC. The approach allows for intra-flow monitoring and flow-based rerouting. A dedicated PM can be used to detect the degradation of a series of high-QoS packets; in this packet protection experiment, an electronic pseudo-BER signal is generated offline by the ParBERT to emulate the output of a PM. An ack is then sent to the transmitter to allow the subsequent packets within the data stream to be rerouted to the protection path. An emerging PM device (*e.g.* an interferometer-based OSNR monitor [34]) can also be used to monitor the flow (as discussed in [35], as well as in the following sections). Depending on the packet's encoded QoS and the measured quality (BER), the CPLD's control logic makes flow-level routing decisions.

5.2.2 Experimental Results

Optical packets with two differing QoS classes are routed through the test-bed. Figure 5.11 provides the optical waveform traces associated with the experiment, including the packets' frame, QoS (high/low priority), and payload; the BER (high or low) is also indicated for each packet. Fast packet switching is confirmed by the packet

sequence, showing the correct routing of all optical messages. The packets correctly emerge at the destination ports that are encoded in their control headers. Packet A is injected in the first timeslot, with a low-QoS/priority encoded in its header. Thus, though it is degraded at the output (indicated as a high BER), the cross-layer control logic allows the packet to be forwarded to the output. It is seen that packet A egresses at the output and a short ack signal appears. Similarly, packet B is injected in the following timeslot with a low-QoS requirement; with a low measured BER, packet B is correctly routed to the output. It is then confirmed that regardless of BER, packets with low-QoS are not discarded by the implemented cross-layer switching scheme. In the third timeslot, packet C is injected in the fabric test-bed using OPS1 with a high-QoS requirement and is denoted as degraded (with high BER). The cross-layer logic detects the presence of a degraded high-priority flow and makes the decision to discard packet C, proactively protecting the stream. Thus, packet C within the high-priority flow and with a high BER is shown to be proactively protected and PPT then reroutes packet C on an alternate protection path, *i.e.* using the parallel network switch. C is retransmitted using OPS2, and, as shown by the waveform traces in Figure 5.11, emerges from the test-bed one timeslot later (due to the different propagation times of OPS1 and OPS2).

At the output of the test-bed, the packets are monitored using an OSA and a high-speed CSA. The CSA is used to capture the traces in Figure 5.11 either at the test-bed input or output. As shown in Figure 5.10, the packet analysis system uses a tunable optical filter that selects one payload wavelength channel at a time for integrity analysis and system performance validation; the performance of all eight payload channels is

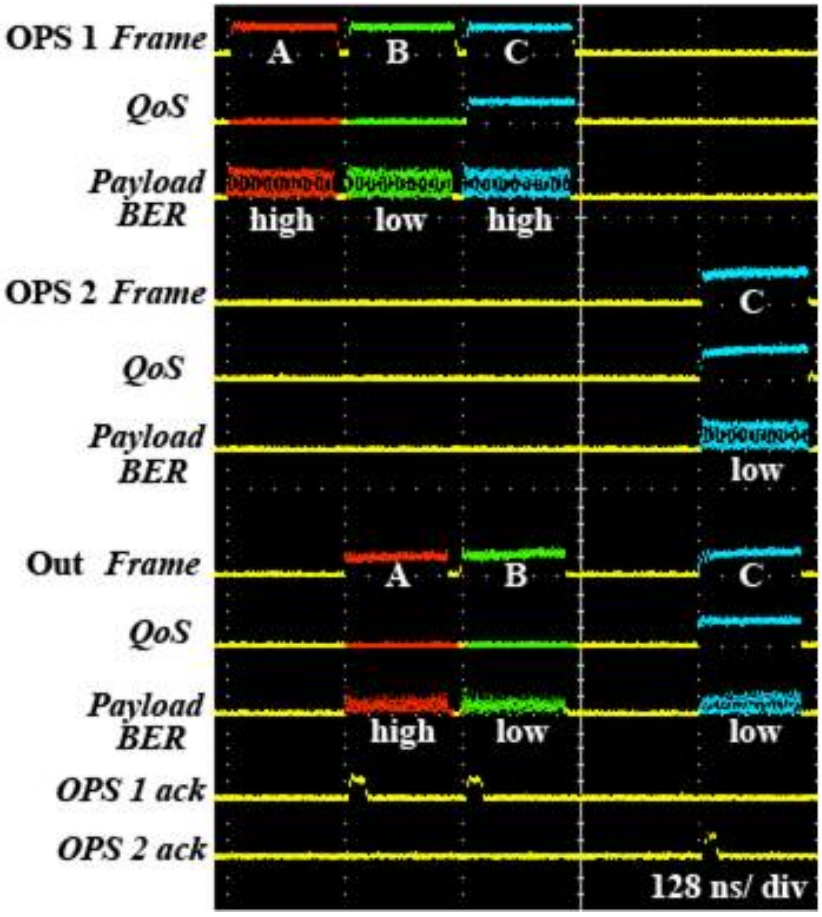

Figure 5.11: PPT Waveforms - Input and output optical waveform traces corresponding to the experimental packet sequence. The colors refer to the different time-division-multiplexed packet streams.

verified by adjusting the filter. The payload channel is passed through an EDFA, another tunable optical filter, and a VOA. The packet is then sent to a DC-coupled 10-Gb/s *p-i-n* photodiode with transimpedance amplifier and limiting amplifier pair (RX). The resulting electrical signals are transmitted to a BERT that is synchronized with the PPG and the packet generation signals. No clock recovery scheme is realized here. The BERT is gated for packet analysis by the ParBERT over 80% of the packet duration; multiple messages from the overall packet sequence were used to determine the resulting BER.

The error-free transmission of the received packet stream is confirmed at the test-bed's output. BERs less than 10^{-12} are measured on all eight payload wavelengths of the received packets. The power penalty performance of the experimental system is then evaluated for all the egressing packets. Figure 5.12 provides the representative sensitivity curves for one payload wavelength channel ($\lambda = 1558.6$ nm). The back-to-back BER curve corresponds to the packet before injection in the router. The through BER curve corresponds to packets that are not proactively dropped by the PPT scheme, and propagate through the three-stage switch fabric as well as the single-SOA stage cross-layer node (*e.g.* packets A and B). A 1.3-dB power penalty taken at a BER of 10^{-9} is measured for this worst case of four SOA hops through the experimental test-bed. Thus, the experimental protection and monitoring design is shown to induce a minimal power penalty. The insets in Figure 5.12 show the 10-Gb/s input and output optical eye diagrams associated with the back-to-back and through packets.

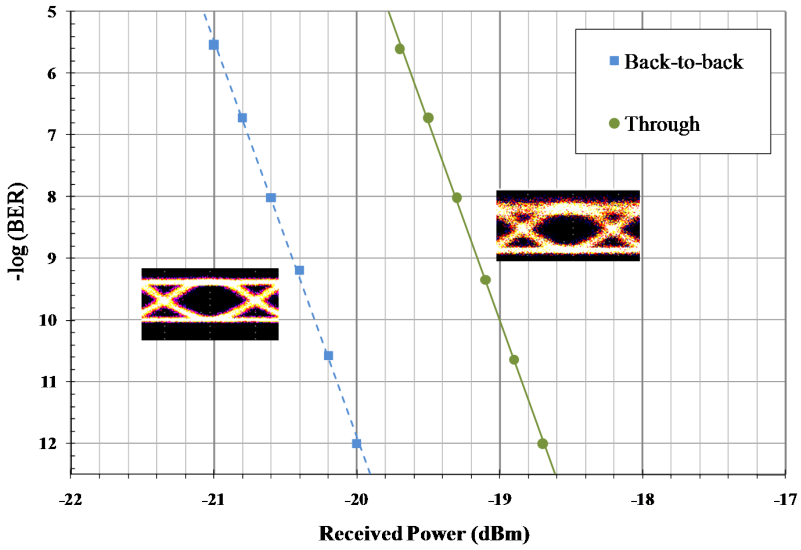

Figure 5.12: PPT Sensitivity Curves - 10-Gb/s BER curves recorded for the back-to-back (depicted by dashed line, unfilled points) and through (depicted by solid line, filled points) cases for the experimental system (for λ=1558.6 nm). Insets depict the input and output optical eye diagrams at 10 Gb/s.

5.2.3 Simulation Exploration

In order to compare the PPT and FRR mechanisms from a systems level, new modules are implemented (Figure 5.13) in the network simulation environment ns-2 [145]. In contemporary network design, state-of-the-art discrete-event packet network simulation tools support a large number of network architectures and protocols, while generally lacking realistic physical-layer models. The implementation of such models in a unified manner is challenging due to the mere variety of possible impairments for optical channels, which may be static in nature or dynamically varying on different time scales. Direct implementations of physical-layer models into packet network simulators are therefore prone to be suitable only for a very limited number of scenarios (*i.e.* modulation formats, impairments, *etc.*) and may become outdated after a while. However, in order to study optical networks in the proposed holistic, cross-layer optimized approach, physical-layer modeling is important if one has to:

- identify the key differences and evaluate the performance of several types of packet networks with respect to physical impairments and dynamic traffic characteristics;
- provide a comparison for different QoS requirement based protocols and new network control efforts such as cross-layer communications.

External physical-layer simulation software (such as Matlab, VPI Transmission Maker, OptSim, STK, *etc.*) or even real-time measurements are used to model physical impairments, resulting in time-dependent BER variations, which are then linked to the packet network simulator.

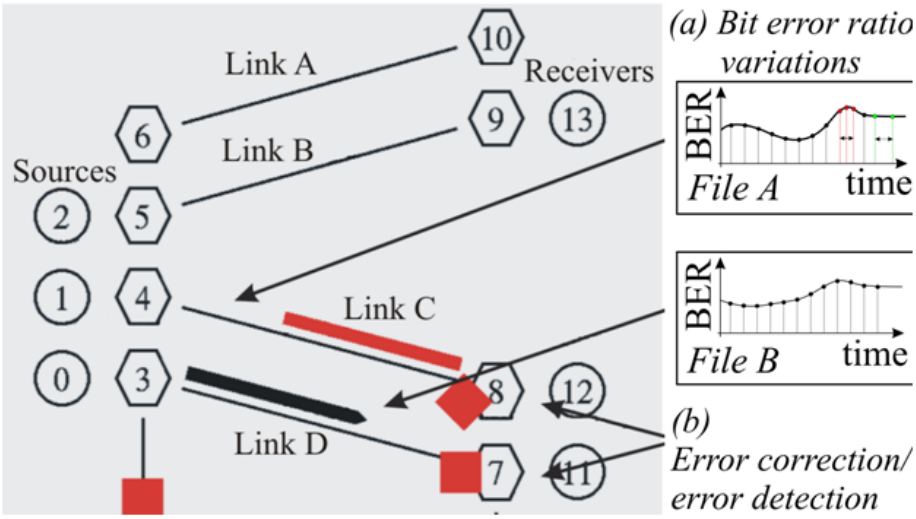

Figure 5.13: Diagrams of ns-2 Modules - The newly implemented ns-2 modules: for (a) intra-packet BER variations, and (b) error correction and error detection.

Using this method, BER variations of varying dynamics can be taken into account, including quasi-static impairments (such as loss or CD), moderately fast impairments (such as PMD [148]), and fast dynamic impairments (such as power transients [149], or nonlinear crosstalk between WDM channels [150].) The BER measurements may be accessed by FEC decoder modules embedded directly in the physical layer.

Two flavors of BER variations are studied here: both packet-by-packet BER variations wherein the BER is constant over the duration of a single data packet, as well as intra-packet BER variations which can potentially lead to burst errors within a single packet. The two types of temporal BER variations of a WDM channel are illustrated in Figure 5.14. The encountered type of BER variations depends on the dynamics of the BER itself, as well as on the length of the data packets and on the underlying network architecture (whether packet-switched or circuit-switched networks

are being considered).

When packets are directly transmitted over an optical infrastructure (such as IP-over-WDM), the majority of physical-layer impairments will be slow relative to the duration of a packet. For example, a packet with an Ethernet MTU of 1500 bytes at a data rate of 10 Gb/s will be approximately 1.2 μs long which is shorter than most dynamic optical impairments. In this case, it suffices to assign a single BER value to a single packet.

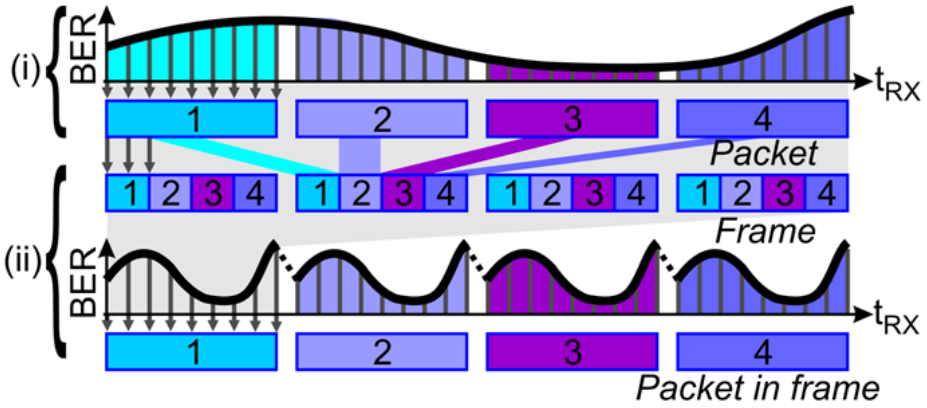

Figure 5.14: ns-2: BER Variations - Diagram indicating the mapping of the BER variations onto bit errors (i) within a packet on a packet-switched link, and (ii) the time compression effect of BER variations when segmenting the packet into multiple TDM frames on a circuit-switched link.

In contrast, intra-packet BER variations may occur in the case when the BER varies quickly relative to the timescale of the packet. This may, *e.g.* , occur because of fast power transients [149] or fast PMD fluctuations. It may also occur for relatively slow fluctuations when the packets are transported on a time-division-multiplexed (TDM) infrastructure (such as IP-over-SONET/SDH/OTN) (Figure 5.15(i)). In this case, the contents of a single packet can be segmented and distributed among many

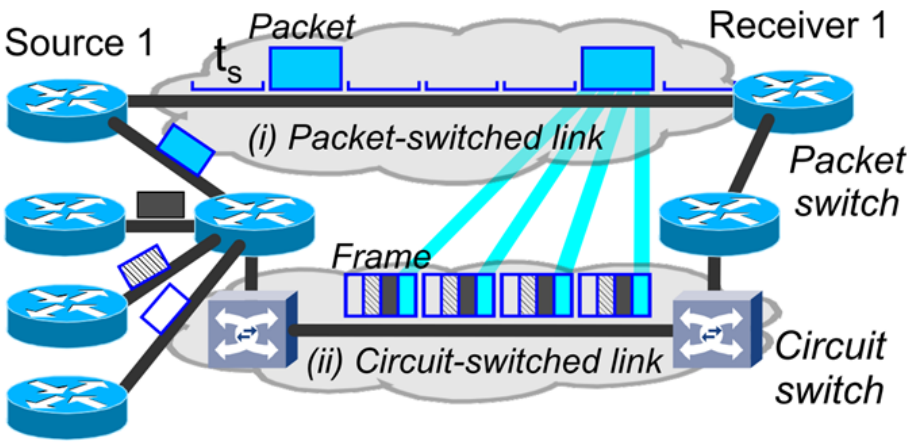

Figure 5.15: ns-2: Packet and Circuit Infrastructures - (*i*) Packet-switched and (*ii*) circuit-switched network architectures showing the difference between packets and frames.

transport frames. When the packets are multiplexed into a frame, time compression and interleaving allow relatively slow physical-layer BER variations to effectively appear as fast fluctuations over the duration of a specific packet (Figure 5.14(ii) and Figure 5.15). For example, an Ethernet jumbo-frame with a MTU of 9000 bytes from a 1-Gb/s client will occupy a timeslot of $t_s = 720$ ns on a 100-GbE packet interface. Using IP-over-OTN, the packet will ultimately occupy $8{\times}9000/10^9 = 72$ μs, which are distributed over 62 OTU4 frames at 112 Gb/s. Here, intra-packet BER variations may be caused by power transients. This issue may be exacerbated for larger packets (such as IPv6 jumbograms with MTUs up to $4{\times}10^9$ bytes) or by protocols with finer granularities than offered by OTN (such as the case with SONET/SDH).

In order to accurately represent these intra-packet BER variations, new ns-2 software modules are realized [146] that calculate the number of bit errors and burst errors per packet according to the BER value stored externally. An additional ns-

2 header was added to each packet that stores information about the transmission and receive times by means of timestamps, the number of bit errors, the number of burst errors, and an error flag. As a packet traverses a link, the bit and burst errors incurred by the packet are calculated according to the transmit and receive times of each bit in the packet and the BER during this time period. For a constant packet-level BER value, bit errors are statistically uniformly distributed within the packet. During periods with increased BER, the probability for burst errors increases for the bits which traverse the link during that time. The calculated bit errors and burst errors within each packet are recorded. Each receiving node then decides whether these errors can be corrected by a FEC device or whether the packet must be dropped.

A new error correction/detection module is also added in ns-2 that allows for the specification of whether a FEC device for error correction is used in the receiving node, whether a cyclic redundancy check (CRC) is performed for error detection, if bit errors should be ignored, or if the packet should be dropped when it contains bit errors. The decision can also account for the required QoS class, as denoted by the modified packet header. Regarding the FEC, the user can specify the maximum number of correctable errors, the maximum number of detectable errors, and the maximum number of consecutive erroneous bits within a packet which may be corrected (*i.e.* the maximum burst error length), in order to account for the different versions of FEC available today. If these values are exceeded by the number of bit errors and burst errors stored in the ns-2 packet header, the message is dropped and/or marked with an error flag. Including the effect of burst errors can be important for certain FEC codes, where an average BER is no longer sufficient to indicate system-level performance;

rather, the statistics of error bursts and the burst error correction capabilities of the deployed FEC are required to quantify performance [151].

The position of the new modules (*i.e.* errmodel_ and errcorr_) within the logical structure of an ns-2 simple transmission link is depicted in Figure 5.16.

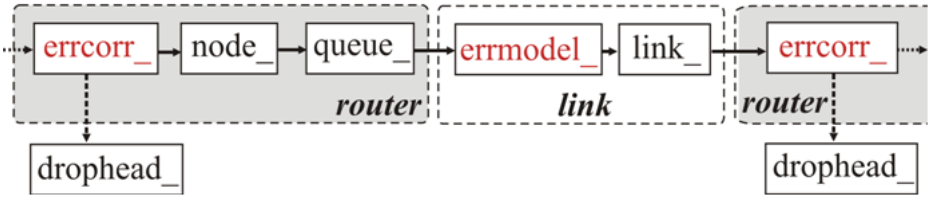

Figure 5.16: ns-2 Transmission Link - Logical structure of ns-2 transmission link including new modules for intra-packet BER variation (errmodel_) and error correction / error detection (errcorr_) (indicated in red font).

5.2.4 Simulation Results

The enhanced ns-2 environment is used to compare the packet loss and throughput performance of the two fast packet protection mechanisms: the PPT scheme and a FEC-based FRR scheme. Flow rerouting occurs for FRR if the resulting BER is greater than FEC's BER_E threshold which is chosen to be $2{\times}10^{-3}$ in this example. The proactive PPT scheme actuates rerouting at a more stringent BER_T threshold (where $BER_T < BER_E$), here set to 10^{-4}. It is assumed that the minimum exchange time for the required control signals between a given receiver and transmitter is limited by the latency time LT. A reasonable quantitative comparison can be made between the PPT and FRR schemes as the data flows are evaluated with a packet granularity. The focus is on a packet-switched core, where a MTU of 1500 bytes from a GbE client would occupy a timeslot of t_s= 120 ns on a 100-GbE interface. Further, a circuit-

switched design is also examined, using the same GbE clients, with a MTU of 1500 bytes and a 100-Gb/s line rate.

The first assumption is a step-like change in the BER (Figure 5.17a, inset (i)) and the number of lost packets n is studied in the network as function of the BER step's rise time. Proactive packet protection (denoted by the green region in Figure 5.17a) offers zero packet loss until the BER step becomes so steep that the time span t_p between BER_E and BER_T is shorter than the latency between transmitter and receiver. For relatively fast BER fluctuations, the number of lost packets using PPT converges to the value $n_t = 2 \cdot \mathrm{LT/ts}$, as LT is the minimum time it takes before the protection mechanism kicks in, *i.e.* the minimum time it takes the receiver to notify the transmitter that the predefined BER threshold was exceeded. Thus, the value of n_t is also the number of lost packets in the case of FRR switching, regardless of the BER variation speed (denoted by the red region in Figure 5.17a).

Figure 5.17a clearly shows the improved packet loss performance of the proposed PPT scheme in the case of relatively slow BER changes.

In Figure 5.17b, the performance savings with respect to retransmitted packets is shown when using PPT as compared to FRR. This savings diminishes with an increasing gradient of the step-like BER degradation and also with increasing LT. Not needing to retransmit packets is of particular advantage in optical packet switched networks, as described in Chapter 2, since optical buffer sizes are small and these buffers are perhaps even nonexistent. As the slope of the BER increases (*i.e.* becomes steeper), the gain vanishes. The reason is that the time span between BER=BER_E and BER=BER_T becomes shorter, thus reducing the benefit of switching to the protection

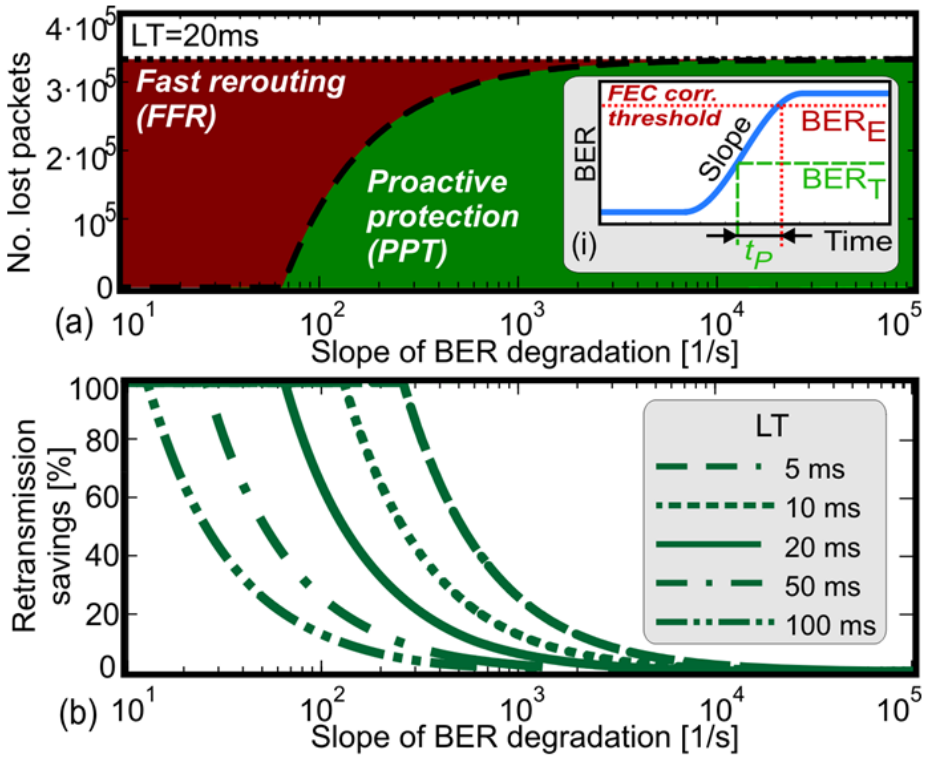

Figure 5.17: ns-2 Results I - Number of lost packets (dotted line refers to the case of FRR, dashed line refers to the case of PPT) in relation to the slope of the BER step (depicted in inset (*i*)). (b) Retransmission savings versus gradient of BER step for varying latencies LT.

path in a proactive way. It is also seen that PPT provides the greatest benefit for short LTs, *i.e.* for small networks. As the latency increases, it takes the receiver longer to notify the transmitter and PPT is less effective than FRR at slower BER variations.

Next, the packet loss rates are studied as a function of the periodicity of sinusoidal BER variations on a logarithmic scale with period Δt (Figure 5.18, inset (i)). PPT offers no packet loss for the situation where the PPT time t_P is greater than the latency. Assuming packet lengths of t_S much less than LT, the two protection schemes perform identically for $t_P + t_E$ =2·LT, with t_E representing the time period during which packet loss occurs (the red region in Figure 5.18, inset (i)). Beyond the point where $t_P + t_E \leq$ 2·LT, the latency of the network is large relative to the impairment dynamics and we see that PPT no longer offers any substantial performance gains over FRR. For both switching mechanisms, the number of lost packets decreases with increasing BER variations, converging to the value n = $2 \cdot t_E \cdot \text{LT} / (\Delta\text{t} \cdot t_S)$. The value t_E/t_S yields the number of lost packets during one peak of BER variations and this number is multiplied by 2·LT/Δt, which is the total number of peaks during which packets are lost before the transmitter is notified by the receiver that the BER threshold has been exceeded and that packets should be rerouted. The slight ripple performance in the packet loss is related to the interplay of LT, t_E, and Δt.

Independent of the protection scheme investigation, the behavior of the packet and circuit switching infrastructures are compared (Figure 5.18a). The solid blue curve in Figure 5.18a corresponds to a circuit-switched core infrastructure. It is found that the two network designs perform similarly until the point where the effective BER dynamics seen by the stretched TDM packets are comparable to the stretched packet

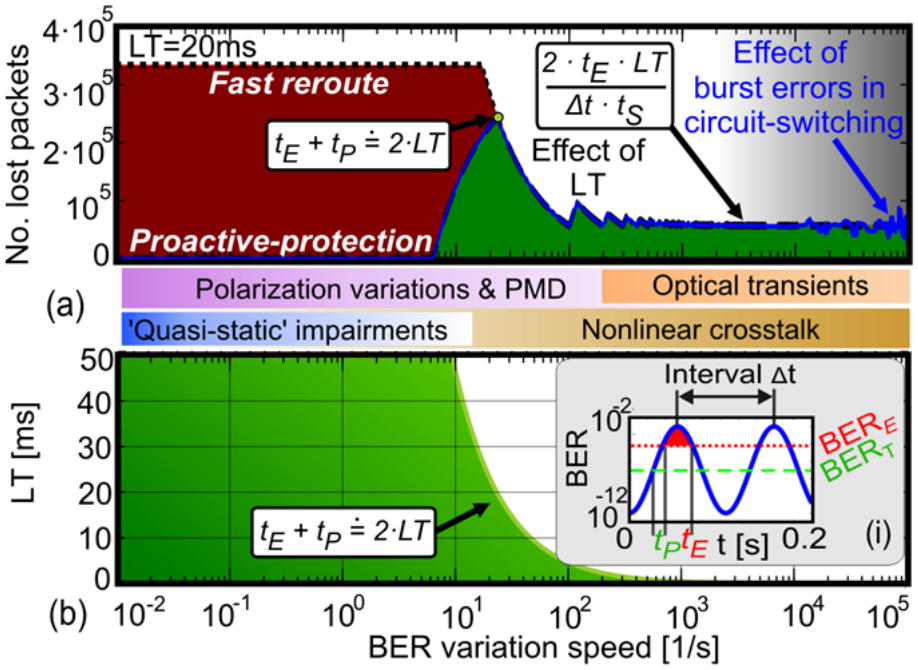

Figure 5.18: ns-2 Results II - (a) Number of lost packets (dotted line refers to the case of FRR, dashed line refers to the case of PPT) in relation to the BER variation speed $1/\Delta$t for sinusoidal BER variations (depicted in inset (i)); (b) Region for which PPT outperforms FRR, in relation to LT and the BER variation speed.

durations (here: $8 \times 1500/109 = 12$ μs). This results in additional variations in packet loss, caused by bursty uncorrectable intra-packet errors.

Using the sinusoidal BER fluctuations, the PPT and FRR protection mechanisms are also compared for various combinations of LT and BER dynamic speeds $1/\Delta$t (Figure 5.18b). The green area under the curve $t_P + t_E = 2 \cdot$LT demonstrates the region where PPT outperforms FRR. Thus, for slow BER variations and larger network sizes, PPT provides a means for achieving substantially improved performance as compared to a FEC-based FRR switching scheme.

Therefore, further confirming the findings in [143], it can be said that fast impairment dynamics require short LTs for PPT to provide an advantage over the standard FRR scheme. In the case of typical optical transport networks, where the latency ranges from 2 ms to 20 ms, PPT is found to be effective against quasi-static impairments and PMD; however, this scheme is likely to fail for dynamic impairments such as fast amplifier power transients. Further, these cross-layer network simulations verify that knowledge of the physical-layer BER variations with varying impairment dynamics is important for accurate studies of packet rerouting schemes.

To close this section: within the scope of cross-layer optimization, this section develops a cross-layer proactive packet protection switching scheme whereby degraded messages within a data flow can be proactively detected with cross-layer logic. This technique is further leveraged in other work (discussed in the following sections). Flow rerouting can thus be triggered by physical-layer signal degradation that is measured on a flow basis. The performance measurements will help enable improved optical flow control for the overall network; the rerouting to an alternate path can be initiated with

minimal latency. Once the rerouting has been established, there are no other additional transients associated with the flow rerouting scheme.

Key cross-layer opto-electronic technologies are demonstrated that support the protection mechanisms, allowing for routing decisions to be optimized dynamically based on the applications' QoS requirements and the signal quality as denoted by the BER. Wavelength-striped optical packets are switched on the fabric test-bed with verified error-free transmission. BERs less than 10^{-12} are obtained and a small 1.3-dB power penalty is induced by the experimental system and the packet protection scheme. Further, ns-2 simulations compare the PPT scheme with FRR mechanisms, showing the support of detailed, granular physical-layer awareness that accounts for fast BER variations. Results show that the packet protection mechanism outperforms FEC-based rerouting for fast impairment dynamics on networks with short round-trip times. New design paradigms will be necessary to support the intensive bandwidth requirements in future networks. This work shows a framework for QoS-aware cross-layer communications that is based on emerging optical technologies, achieving significant performance gains.

5.3 Quality-of-Service-Based Multicasting

This section further expands on both the MPMA packet-multicasting work (described in Chapter 4), as well as the notion of QoS-based routing (as motivated in Chapter 2), by incorporating these functionalities in an integrated framework. To this end, QoS-aware cross-layer multicasting for optical packet switching fabrics is explored [152]. A numerical simulation exploration of the cross-layer algorithms is performed, in

conjunction with an experimental demonstration on the optical switching test-bed with 10×10-Gb/s wavelength-striped messages.

The envisioned design (Figure 2.6) will allow future networks to extract introspective OPM measurements directly from the optical layer to globally optimize performance. These cross-layer schemes and routing protocols must also invoke QoS requirements on the optical layer to provide application-specified QoS guarantees to data flows. Ultimately, these optimized algorithms must both provision for the data's QoS as well as account for the physical-layer performance and impairments. Deployed OPS fabrics must support high-bandwidth multiwavelength packet streams between line cards and must achieve an advanced level of programmability to transparently route WDM packets entirely in the optical domain. Furthermore, as outlined previously, broadband packet multicasting is a significant application that may leverage the greater functionality and programmable flexibility offered by the optical physical layer.

Specifically, broadband QoS-based packet multicasting constitutes an important functionality for future OPS fabrics. Notably, for bandwidth and latency sensitive applications (such as real-time collaboration or teleconferencing), high-QoS packet transmission may be leveraged to provide an high-quality communication link for a fixed, predetermined duration. This section explores a cross-layer enabled platform whereby a packet multicasting operation is realized accounting for both the data stream's QoS and physical-layer degradation. The concept of cross-layer QoS-aware multicasting is investigated both in simulation and with a test-bed demonstration. A simulation-based comparative analysis is first provided between shortest distance and minimum hop routing algorithms using the NSF network topology, showing the latency,

hop count, and packet loss performance when accounting for varying levels of QoS. The following experimental demonstration shows the OPS fabric implemented within one node of NSF network, validating the error-free operation of cross-layer QoS-based multicasting with BERs less than 10^{-12} and a power penalty of 2 dB.

5.3.1 Simulation Exploration

The proposed routing algorithms for QoS-aware packet multicasting are first investigated through simulations. One-way based signaling is assumed to minimize the end-to-end packet transmission latency. The 14-node NSF network topology (Figure 5.19) is numerically simulated using a global control plane to track each node's QoS performance. A centralized routing and wavelength assignment (RWA) scheme is realized, and the intelligence of the control plane architecture is built to have the best path chosen based on QoS parameters. Messages are assumed wavelength-striped, using ten wavelength channels each at 10 Gb/s.

Packets are simulated as discrete events. These packets follow a Poisson arrival rate and depart with exponential service times. On an arrival event, each packet is assigned to a request and then routed based on a minimum distance routing (MDR) or a minimum hop routing (MHR) algorithm. The necessary QoS parameters are retrieved from the application layer, and the packet is routed based on this information exchange. Optical packets reaching the destination ensure that the threshold requirements imposed by the application layer are met [57, 118]. QoS information is embedded in the control signal and is updated as the packet propagates towards the destination. At each network node, the QoS information of the packet that is being routed is compared

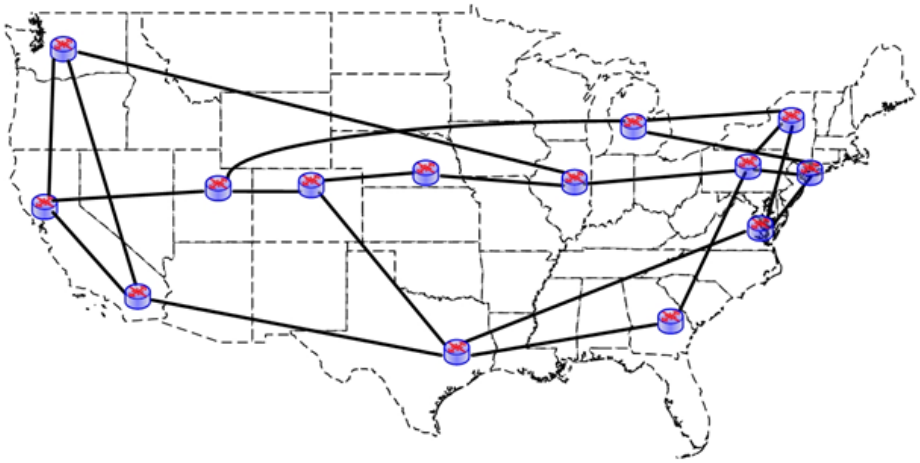

Figure 5.19: QoS Aware Multicasting: NSF topology - Diagram of the 14-node NSF network topology with bidirectional links between the nodes, each carrying ten wavelengths.

with the threshold requirement of the application that maps to the packet. If the QoS parameters are violated, the packet is dropped or rerouted on an alternate path if available.

An intelligent and efficient control plane acts as a middleware between application and optical layer; Figure 5.20 depicts the envisioned control and management layer. It is assumed that the transmission of the optical packet is followed after a specific offset time. Based on the control plane decision, the optical packet is routed on the established optical lightpath.

The QoS information in the control packet consists of optical signal's BER, latency, priority, and the reliability of the link on which the packet is traversing. At each node between the source and destination, the online computation of the QoS parameters associated with each packet is performed and the multicasting operation is initiated as

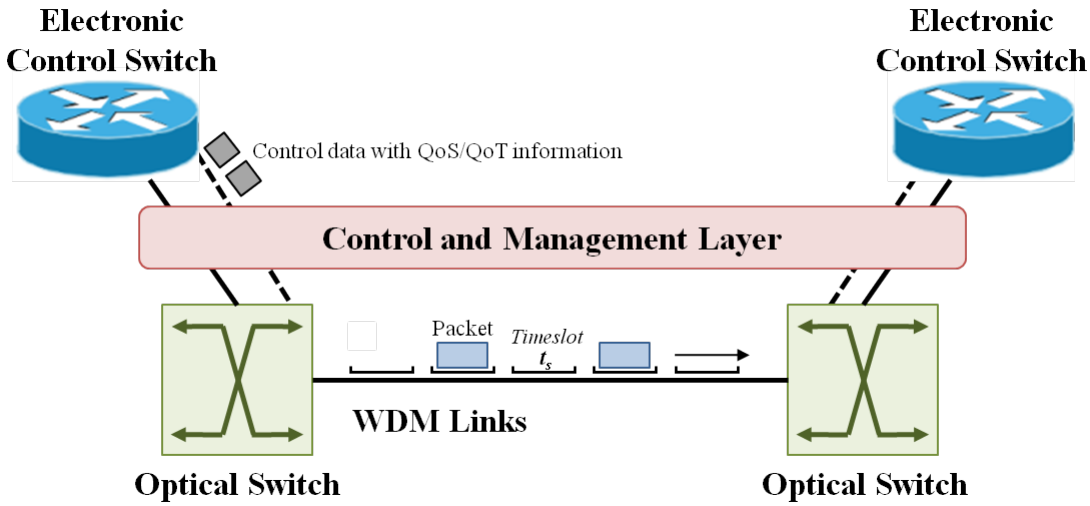

Figure 5.20: Control and Management Layer - Block diagram of the proposed control and management layer existing above the optical physical layer.

required.

In these simulations, the BER is estimated based on the OSNR. Optical signal power degradation occurs due to the device components and noise accumulated in the optical amplifiers and switches. The parameters and threshold values used in the BER estimator are shown in Table 5.1. Since BER is a nonlinear function, the link's noise factor is computed. The overall noise factor of the link is computed as a product of the noise factors of the links. The overall latency for the optical packet is sum of the individual latencies in a link. The reliability of the wavelength switch is based on the regeneration of the optical signal, switch downtime, and path restoration time. The priority information enables the search for possible alternate paths for routing the optical packets.

The performance of the proposed QoS-aware cross-layer multicasting is simulated

Table 5.1: Parameters used in the QoS-aware packet multicasting simulations.

Parameter	Value
Number of packets	10^6
BER	10^{-9}
Latency	1 ms
Optical Power	-10 dBm
Inline Amplifier Gain	14 dB
Switch Crosstalk Ratio	25 dB
Wavelength Spacing	2.8 nm
Starting Wavelength	1537.4 nm

on the NSF network topology with the distances scaled down by a factor of ten. The present network does not deploy regenerators at the node's switching fabrics, and hence the scaling was used here. The performance of the NSF topology is then evaluated with respect to packet loss, execution time for the routing algorithm, average latency of the packet, and hop count taking into account the MDR and MHR algorithms.

The independent variable (*i.e.* the x axis in the plots) in the following analyses is the offered network load in Erlang, defined as the ratio of arrival rate to the departure rate. In Figure 5.21, it is observed that MHR offers lower loss compared to MDR at low network loads. This indicates that optical packets routed based on hop count show a higher probability of successfully guaranteeing the QoS imposed by the upper application layer. The average latency (shown in Figure 5.22) of MHR is higher compared to the distance routing, which may not be a concern from the application layer's perspective, as long as the threshold requirement for latency is satisfied. Thus, the routing layer can choose to adopt a hop based routing at lower network loads. As the network loads increase, the packet loss for both routing algorithms converges in

Figure 5.21. In order to have optimal performance, the application layer can instruct the routing layer to switch to distance routing at higher network loads. Thus, this cross-layer design helps to achieve trade-offs and provide the necessary QoS.

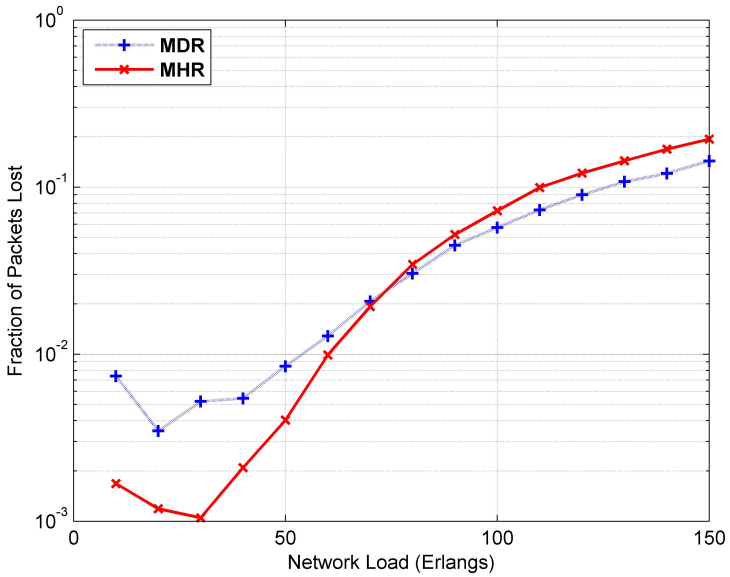

Figure 5.21: QoS-Aware Multicasting: Blocking Performance - Packet blocking performance of the scaled NSF network, for the MDR and MHR schemes, showing the fractions of lost packets.

The average hop count for the two routing methods is also compared. It is obvious that the resulting average hop count for the MHR should be lower than MDR (Figure 5.23). A decrease in these values at higher loads indicates that providing QoS for optical packets that traverse longer hops are more problematic.

Figure 5.24 shows the execution time (in hours) required for performing the simulations on a machine with a 2.33-GHz Quad Core Xeon processor and 8 GB of RAM. The processor also utilizes Hyper-Threading Technology; therefore, these

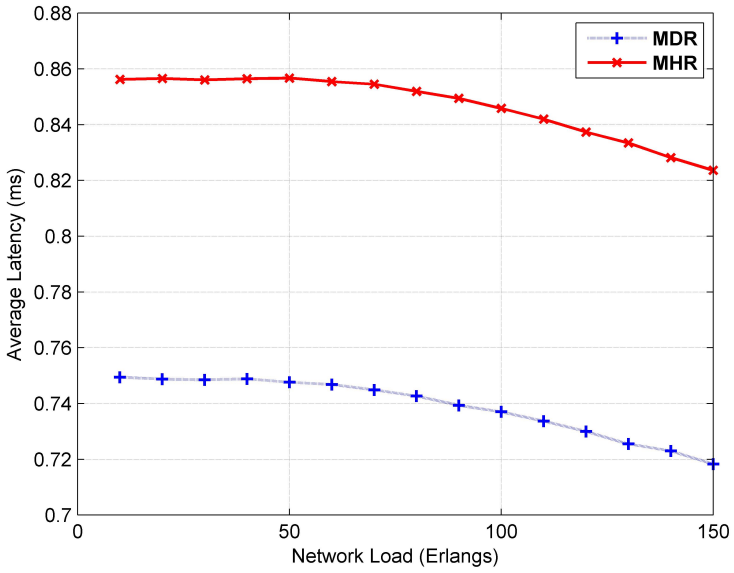

Figure 5.22: QoS-Aware Multicasting: Latency Performance - Latency performance (in ms) of the scaled NSF network.

simulations were able to use eight simultaneous threads while performing and analyzing the discrete events.

5.3.2 Experimental Demonstration

The QoS-aware broadband packet multicasting operation is experimentally demonstrated on the multicast-capable OPS fabric test-bed (Figure 4.20). The test-bed represents the optical switching fabric deployed within one node of the NSF topology described above. The test-bed is implemented using ten non-blocking 2×2 PSEs. The multicast-capable fabric architecture (MPMA, as outlined previously) is realized with a multistage design: a subset of the fabric stages is used for packet routing (PaR) and a subset of stages is used for packet multicasting (PaM). The two sets of stages

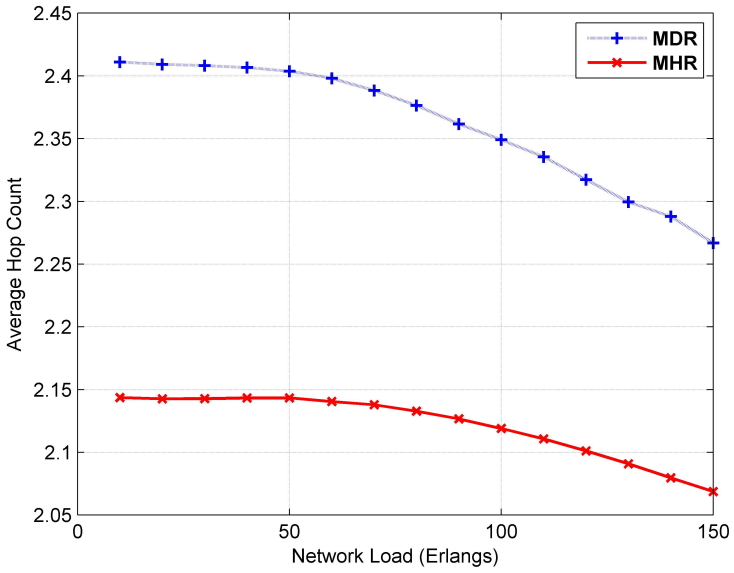

Figure 5.23: QoS-Aware Multicasting: Hop Count - Simulated performance of the scaled NSF network, in hop counts for the MDR and MHR algorithms.

have distinct distributed control logic that depend on the packet's recovered header information.

With the goal of OPM extraction and the implementation of a cross-layer platform, a QoS-aware SOA-based receiver node design is realized [115] whereby the QoT performance of optical packets may be monitored in real-time. This switching scheme is triggered on both the per-packet QoS/priority and signal degradation (represented here as BER threshold). Packets with high-QoS/high-BER are detected by the cross-layer-enabled receiver and rerouted on an alternate path to minimize packet loss. The pseudo-OPM/BER cross-layer signal is generated offline in place of an OPM, though it has also been shown that a real-time packet-based OSNR monitor may be used.

The QoS-aware packet multicasting operation is validated on an implemented 4×4

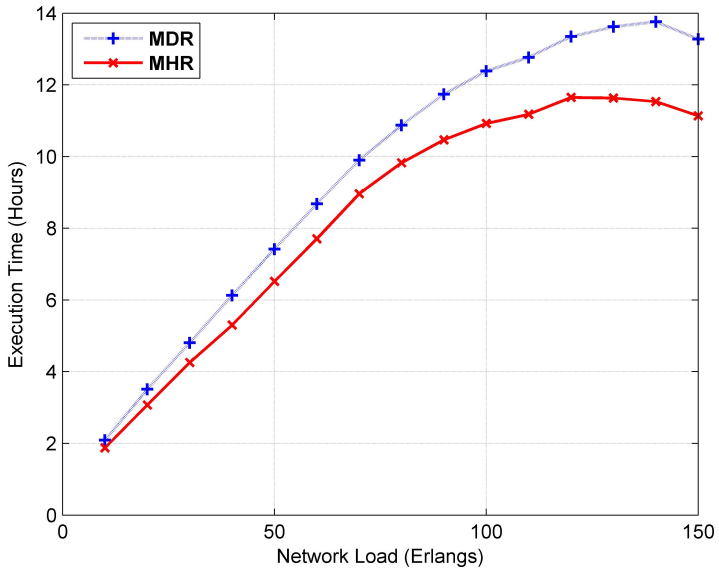

Figure 5.24: QoS-Aware Multicasting: Execution Time - Total execution time for the MDR and MHR schemes, in hours.

optical fabric test-bed with two PaR stages and three PaM stages. The wavelength-striped packets are 120-ns long with a 10×10-Gb/s format, corresponding to the size of an Ethernet MTU. The 1500-byte payload packets are modulated by one $LiNbO_3$ modulator with a 2^7-1 PRBS in NRZ-OOK format, with the payload wavelengths ranging from 1537.4 nm to 1564.0 nm. Figure 5.25 depicts the pattern of optical packets injected in the multicast-capable fabric with two distinct QoS levels (high and low priority). The QoS and packet signal quality are assessed and a proactive real-time decision is made to forward or reroute the message on a protection path.

At the output of the fabric, the cross-layer receiver demonstrates that QoS-based packet multicasting can be achieved with error-free operation. The optical packet is filtered to extract one discrete payload channel, amplified using an EDFA, filtered again

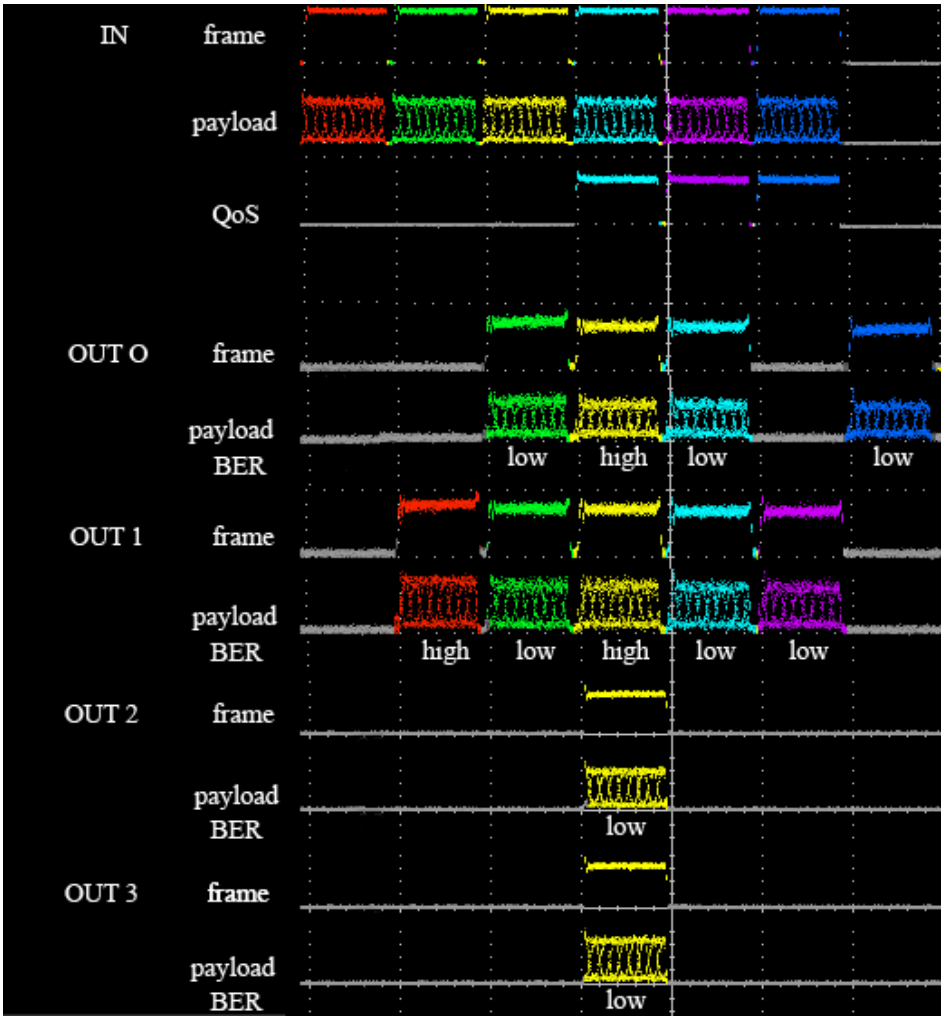

Figure 5.25: QoS-Aware Multicasting: Experimental Traces - Experimental optical waveform traces corresponding to the QoS aware packet multicasting operation, showing the injected and egressing traffic.

to remove the ASE from the EDFA, then transmitted to a VOA. The output of the VOA is connected to a 10-Gb/s *p-i-n* receiver with TIA and LA pair. Bit-error-rate measurements using a BERT confirm that BERs less than 10^{-12} are attained on all ten payload wavelength channels at the fabric output. Sensitivity curves corresponding to the output of the five-stage fabric and cross-layer node show a complete system power penalty of 2 dB (Figure 5.26).

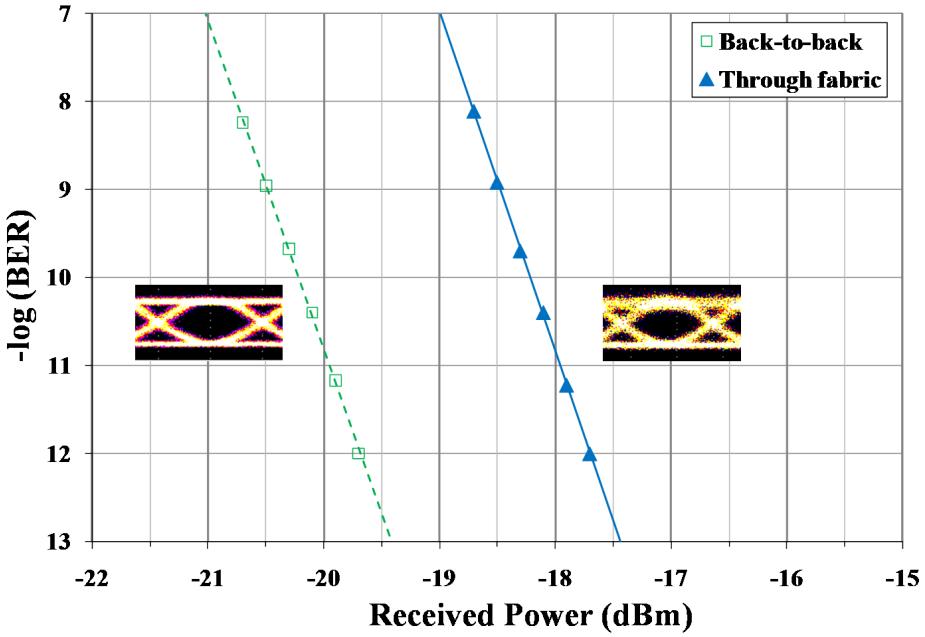

Figure 5.26: QoS-Aware Multicasting: Sensitivity Curves - 10-Gb/s BER curves associated with the back-to-back measurements (green, unfilled points) and the through case (blue, filled points). The 10-Gb/s optical eye diagrams are shown as insets (input: left, output: right).

This work confirms that wavelength-striped packet multicasting can be realized incorporating physical-layer access in a cross-layer-optimized approach. Numerical results and an experimental demonstration on an OPS fabric test-bed show that packet

multicasting can be performed based on both the packet's QoS and signal degradation.

5.4 Packet-Scale Performance Monitoring: An Overview

In order to truly showcase the advantages of this cross-layer framework (as depicted in Figure 2.6), the network must possess deep introspective access to the optical physical layer by means of fast performance monitoring. This is especially true with the possible requirement on next-generation infrastructures to support all-optical transmission at high modulation bit rates. Further, with advancements in Generalized Multi Protocol Label Switching (GMPLS), switching IP packets directly in the optical domain is becoming an attractive option. To ensure reliable, robust high-speed optical data links, these data-centric networks will require advanced OPM capabilities to dynamically measure optical signal degradations in real-time. The need for OPM in future networks and systems has been driven by collaborators such as Willner *et al.* [36, 153], as well as by other leading researchers such as Kilper *et al.* [37]. This comprises a potentially essential functionality for high-capacity networks and will enable the monitoring and isolation of physical-layer impairments, in addition to the fast evaluation of the optical QoT of the transmitted data signals. These metrics can then provide a means of feedback to higher network layers or a control plane to optimize routing [20]. Numerous optical parameters can be monitored, including signal power, wavelength, OSNR [34, 154], CD [155, 156], and PMD [157].

The notion of physical-layer-impairment-aware RWA algorithms has been explored by many researchers (*e.g.* [53, 59, 158], among others; see also Chapter 2), which

is specifically important in the case of transparent all-optical networks. Most of these investigations have leaned toward the simulation side, through the development of network algorithms and routing schemes that can incorporate physical-layer impairments. Here, the author has chosen to take a more experimental approach to these cross-layer techniques and focus on the implementation of the required packet-level OPM and/or PM modules that will be necessary to realize these algorithms in a practical way.

This work envisions performance monitoring within OPS fabrics to help enable an agile network that can independently and holistically isolate degradations and reroute optical messages accounting for impairments. Several researchers have been involved in similar work [159, 160, 161, 162]. The ability to realize fast, real-time performance monitoring of the physical layer allows for these integrated packet-timescale OPM subsystems to efficiently execute routing. Using the proposed cross-layer optimized networking environment, accounting for applications' QoS will help create QoS-aware protocols within a dynamically adaptable network.

The following two sections cover two major lines of work regarding the development of fast performance monitors for OPS fabrics. The first is a demonstration of fast OSNR monitoring in the OPS test-bed (in collaboration with Willner *et al.* at the University of Southern California (USC)) and the second comprises some initial first steps to performing fast BER extrapolations by means of real-time burst sampling (in collaboration with Jalali *et al.* at the University of California, Los Angeles (UCLA)). This work is an active and ongoing research activity within the CIAN initiative, and expansions of this work will likely extend long beyond the writing of this dissertation.

5.5 Optical-Signal-to-Noise Ratio Monitoring

This subsequent section outlines the experimental demonstration of OSNR monitoring of 10-Gb/s NRZ-OOK-modulated optical packets in a cross-layer QoS-aware PPT switching scheme [115, 143], allowing for OPM modules to feedback to upper network layers for packet rerouting and protection [35]. The scheme is implemented on the OPS fabric test-bed, with a realized packet-level OSNR monitor [34] to actuate a rerouting of high-QoS/priority optical messages upon measuring a degraded OSNR. The scheme aims to reduce the penalty associated with packet retransmission of critical, high-priority data flows. The OSNR is monitored using a 1/4-bit Mach-Zehnder (MZ) delay-line interferometer (DLI) implemented in conjunction with power meters and a FPGA. The fast OSNR monitoring system is realized in a cross-layer enabled OPS fabric test-bed to allow the packet protection mechanism to dynamically detect and reroute degraded high-QoS optical messages. Alternatively, degraded messages can be dropped and regenerated to be forwarded to subsequent network nodes.

Fast OPM allows OPS fabrics to dynamically detect and recover from impairments [126, 160], with potential path restoration capabilities [163]. Specifically, OSNR monitors that exhibit fast responses and that can measure the OSNR on a packet-by-packet basis may be a valuable tool for OPS fabrics to ensure reliable links. Here, the fast OSNR monitor measures the performance of 10-Gb/s OOK data streams and facilitates a QoS-aware protection scheme similar to previous demonstrations with a pseudo-BER signal. 8×10-Gb/s multiwavelength packets are shown routed through the OPS fabric test-bed, detected and assessed by the OSNR monitoring system, and rerouted based on the packet-encoded QoS/priority and QoT/OSNR measurement

signals. The wavelength-striped optical messages that egress from the output of the test-bed attain error-free performance with BERs less than 10^{-12}. A power penalty less than 2 dB is obtained using a three-stage optical switching fabric and the packet-timescale optical performance monitoring system.

5.5.1 Cross-Layer Packet Protection Scheme

As in [115] (and above), the proactive packet protection switching scheme uses OPM measurements as an indication of degraded data streams. The switching mechanism is also triggered by the QoS class encoded in the optical packet. A degraded signal quality is detected (here, a low OSNR is measured) and the protocol sets a predefined performance threshold below which rerouting is actuated to prevent the loss of important data streams. The degraded optical packets making up a high-QoS data flow are discarded and a cross-layer control signal is backwards propagated to the source, which can then proactively switch and reroute the data stream on a parallel protection path. The QoS-aware nature of the protocol allows data flows with high-QoS, low-OSNR optical packets to be proactively identified and rerouted on the protection path, while low-QoS (regardless of OSNR) and high-QoS, high-OSNR messages are forwarded to the destination.

The packet protection mechanism may be experimentally implemented in several ways. The OPS switching fabric itself could accept feedback signals from a control plane and/or directly from the OPM devices, or a separate custom switch could be used at the receiving end to forward or discard the message. Figure 5.27 (a slight modification of Figure 5.9) provides a diagram of a potential network node architecture

for the physical layer. Here, a custom switch design is used to realize a new cross-layer receiving node (Figure 5.28), which uses a dedicated OPM that measures the optical signal quality in real-time. At the output of the switching fabric, the optical packets are sent to the cross-layer node such that the optical signal's QoT can be calculated on a packet-by-packet basis. Thus, both the message-level signal impairments and packet-encoded QoS classes can act as inputs in the network routing decision and enable impairment-aware packet switching protocols.

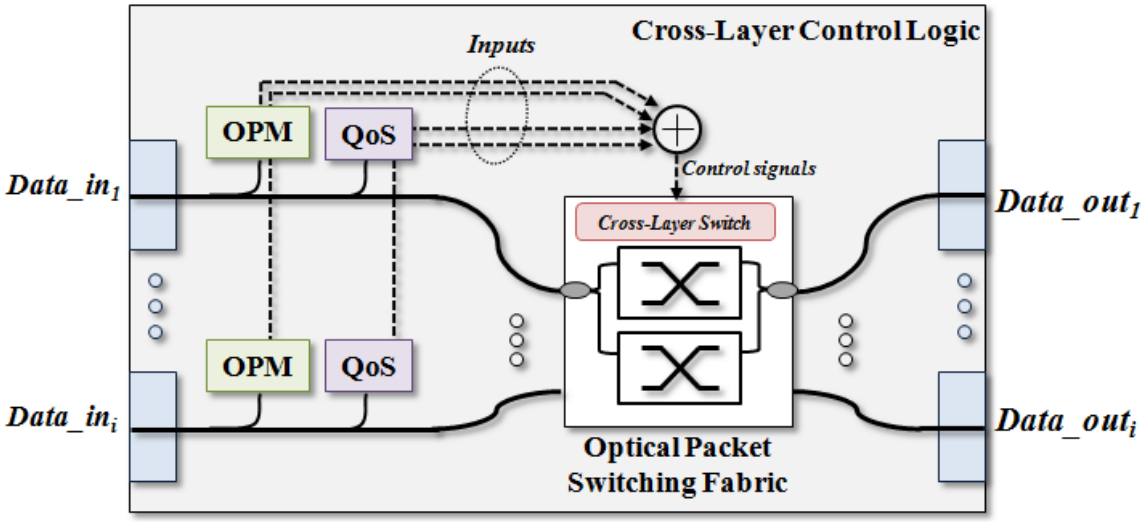

Figure 5.27: OSNR Monitoring Network Node - Architecture of a possible physical-layer implementation of one optical network node incorporating the OPS fabric and OPM modules.

5.5.2 Fast OSNR Monitoring System

The OPM is realized as a fast OSNR monitor, which measures the OSNR on a message basis to ultimately allow packet-level control and rerouting of a data stream using the packet protection switching protocol. The OSNR is assessed using a 1/4-bit MZ

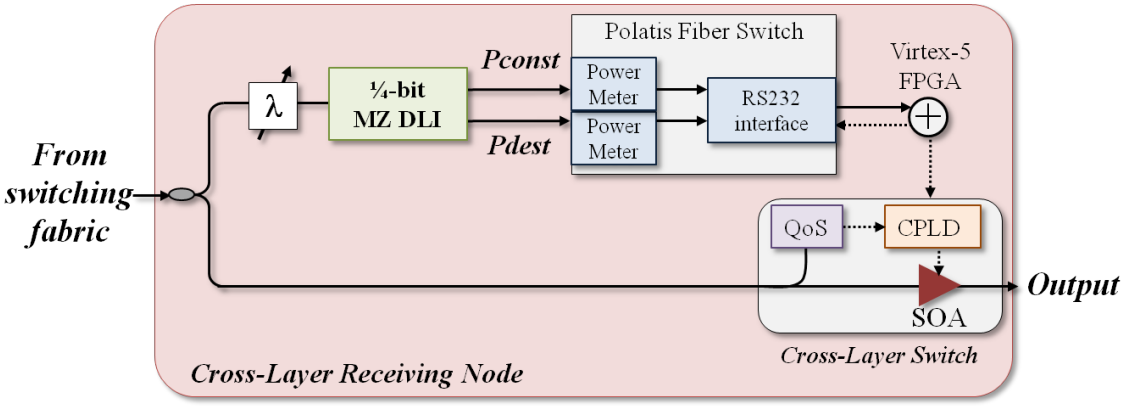

Figure 5.28: OSNR Monitoring System - Experimental Setup of cross-layer receiver node used in this experiment.

DLI, designed to support several modulation formats at 10 Gb/s [34]. The OSNR monitoring method is independent of other physical-layer impairments, such as CD and PMD. The two discrete ports of the DLI provide constructive (P_{const}) and destructive (P_{dest}) interference, respectively. At the output of the 1/4-bit DLI, the constant phase relationship during a single bit yields constructive interference over 3/4 of the bit period. The signal's OSNR is proportional to the ratio of P_{const} divided by P_{dest}. Since the phase relationship between consecutive bits is not crucial to this monitoring method, multiple modulation formats can be supported; here, a NRZ-OOK modulation format is used. The 1/4-bit DLI has a free spectral range (FSR) that is four times the bit rate (here, 10 Gb/s), thus the majority of the power is transmitted to the constructive port. The noise signal is evenly distributed between the two output ports. With a decreasing measured OSNR value, P_{dest} increases greater than P_{const} as a result of the random noise.

In order to monitor the OSNR at a packet timescale, the OSNR monitor uses the

DLI in conjunction with a high-speed FPGA. At the output of the switching fabric, the custom cross-layer receiving node filters a portion of one 10-Gb/s payload channel comprising the egressing wavelength-striped optical packet and transmits the filtered signal to the DLI. The values of P_{const} and P_{dest} are then determined from the two output ports of the DLI, each of which is connected to a power meter. The FPGA obtains the two power values at the packet rate and the high-speed logic performs the online processing to evaluate the P_{const} / P_{dest} ratio to determine the packet's OSNR on a packet timescale. Within the FPGA, the calculated OSNR value is then compared to a performance threshold. If the minimum value is met, the FPGA generates an electronic gating signal that is the length of the packet and transmits the pulse to a SOA-based cross-layer switch that is controlled by its own CPLD. The switch simultaneously extracts the QoS from the wavelength-striped packet's header using a fixed wavelength filter and optical receiver. Using both the packet's QoS (*i.e.* the designated high or low priority) and OPM-based QoT signal from the FPGA, the CPLD ultimately makes a routing decision to switch or discard packets as per the proactive protection scheme. This is accomplished by either gating on or off the separate cross-layer SOA. Optical packets may be forwarded to the final destination port, or discarded and rerouted on the parallel path.

5.5.3 Experimental Setup

To demonstrate the packet protection switching scheme, the OSNR monitoring system is implemented on a multi-terabit capacity 4×4 cross-layer enabled OPS fabric test-bed [26]. The test-bed is based on the multicast-capable PSaD architecture, using two

parallel OPS fabric entities to allow a path diversity of two between the source and destination nodes [128] (*i.e.* the fabric design offers two distinct and independent paths between any of the source ports to any of the destination output ports). The test-bed here uses two parallel optical packet switches (a three-stage switch in conjunction with a two-stage switch) to provide a main routing lightpath and an alternate protection route. Messages are routed across each optical packet switch entity based on the wavelength headers encoded in the individual packet's structure. The ack procotol is realized in this experiment. The aforementioned cross-layer receiver design is realized at the output of the switching fabric test-bed, incorporating the OSNR monitoring system. Depending on the encoded QoS class in the header, packets with degraded measured OSNRs are intentionally discarded by the cross-layer receiver to suppress the ack pulse and thus trigger packet rerouting.

To demonstrate the packet routing and protection functionalities, a predetermined experimental pattern of 8×10-Gb/s wavelength-striped optical packets is generated and injected into one port of the fabric test-bed (Figure 5.29); the packets have differing high and low OSNR values. At the input of the test-bed, the wavelength-striped optical packets are generated as per the following setup. The payload wavelength channels are created using eight discrete CW-DFB lasers. The DFBs range from 1540.1 nm to 1558.3 nm, to explicitly showcase the broadband transparency of the optical switching design. The minimum spacing between two adjacent payload channels is 100 GHz (0.8 nm), to demonstrate that no crosstalk is shown between neighboring channels. All eight channels are passively multiplexed onto a single-mode fiber using an optical coupler and are concurrently modulated using a single $LiNbO_3$ amplitude modulator that is

electrically driven by a high-speed PPG. The PPG generates a 10-Gb/s 2^7-1 PRBS in a NRZ-OOK format, modulating all eight channels simultaneously. The resulting multiwavelength payload information is then decorrelated using 25 km of SMF-28 and passively multiplexed into two identical streams to create the dedicated high-OSNR and low-OSNR data streams that will be required for experimentally monitoring the OSNR at the output of the fabric. As in Figure 5.29, the low-OSNR data streams in the input pattern are generated by degrading the OSNR of the low-OSNR stream using an optical attenuator (set to attenuate by 8 dB), followed by a separate SOA from Kamelian. The SOA amplifies the signal to the original value while concurrently providing a certain amount of ASE to yield sufficiently degraded OSNR streams. The dedicated high-OSNR data stream is not transmitted through this OSNR-degradation setup (of an optical attenuator and SOA), thus resulting in sufficiently high OSNR for this flow. In this demonstration, the threshold for the high and low OSNR classes is 5 dB.

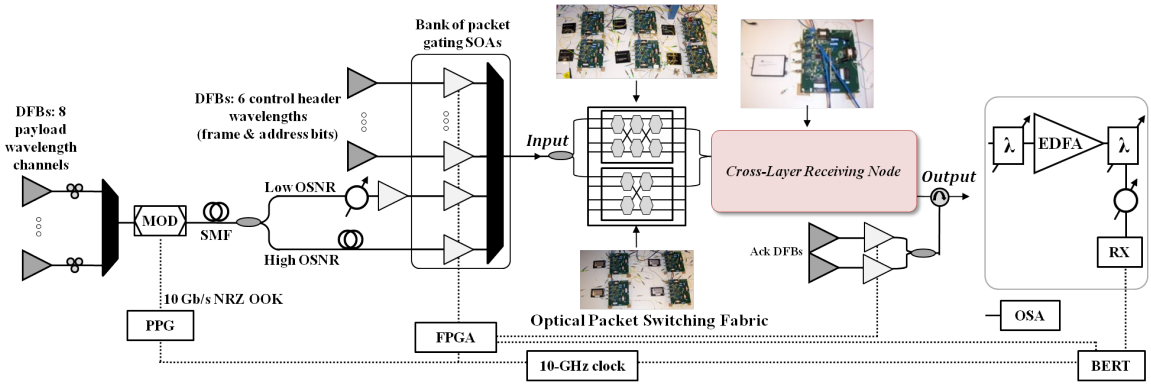

Figure 5.29: OSNR Monitoring Experimental Setup - Experimental setup of the fast OSNR monitoring system with OPS fabric test-bed. The figure depicts the generation of the high- and low-OSNR optical packets, photographs of the optical switching fabric and cross-layer node, and the setup of the packet-analysis system.

The control header signals are generated separately using six other CW-DFB lasers at the required frame, address, and QoS wavelengths for routing. Each of the control header bits, as well as the high-OSNR and low-OSNR payload data streams, are then transmitted to separate external SOA devices which gate the continuous streams into discrete wavelength-striped optical packets for the experimental demonstration. The gating SOAs are driven by a fast FPGA that is synchronized to the PPG and acts as an electrical pulse generator to manage the experimental test-bed addressing and packet gating. The electrical outputs of the FPGA denote a pre-programmed sequence of optical packets that were chosen explicitly for this experiment. The source for the optical ack pulses is a separate DFB laser at 1541.1 nm, which is transmitted to a SOA for gating by the FPGA. The appropriate gating SOAs' outputs are then multiplexed together using a passive optical coupler.

Thus, the created sequence of 8×10-Gb/s wavelength-striped optical packets contains: a constant one-bit frame signal; the appropriate optical address encoding to route the optical messages transparently end-to-end on the test-bed using both parallel packet switches; a QoS control header denoting the high or low QoS class designated to the optical packet; and a combination of high and low OSNR multiwavelength payloads, with the data modulated at 10-Gb/s NRZ-OOK on eight frequency channels. The external SOAs are gated by the FPGA in such a way that a single pattern of packets will have two differing OSNRs; this is realized by ensuring that the gating pulses for the high-OSNR and low-OSNR data streams do not overlap. This yields a pattern sequence of both high-OSNR messages and low-OSNR messages, with two different QoS classes, which is injected in one active port of the switching fabric test-bed to

demonstrate the protection scheme.

Similar to [115], the QoS class is directly encoded as one of the control headers, and is set constant over the length of the packet; a high-QoS class is denoted by a high control bit, while a low-QoS class is represented by a low control bit. To accommodate the sampling timescales required by the cross-layer receiver, OSNR measurement, and FPGA processing, the system here supports 18-ms packet lengths within 20-ms duration timeslots. The packet durations are determined by the recovery times of this initial, discrete-component experimental implementation of the fast OSNR monitor and are limited primarily by the processing speeds of the power meter equipment and of its communication interface. Future integrated setups will allow for even more rapid OSNR measurements.

To realize the faster monitoring of the OSNR, the cross-layer receiving node is implemented at the output of the fabric test-bed using the setup in Figure 5.29. The cross-layer node allows for a packet-level OSNR monitor that can dynamically monitor fabric-egressing messages and can initiate a rerouting of degraded-OSNR high-QoS packets. Here, the cross-layer receiver is realized at one of the test-bed ports for demonstration purposes; in future scale implementations, a similar design is envisioned at each output port to fully exploit the protection routing functionalities for all egressing optical packets. The receiver uses the aforementioned interferometric-based OSNR monitor, implemented with a high-speed FPGA and SOA-based switch. At the receiving end of the fabric test-bed, a portion of the wavelength-striped packet is filtered using a JDSU TB9 tunable grating filter with an optical bandwidth of 0.22 nm. The tunable filter is set to select any of the 10-Gb/s payload channels in the egressing

optical packet. The OSNR is directly dependent on the effective bandwidth of the wavelength filter, as it determines the noise equivalent bandwidth of the monitored 10-Gb/s channel [126]. The system here monitors 10-Gb/s NRZ-OOK signals; with the above filter bandwidth and supported modulation format, a P_{const} / P_{dest} ratio of 7 dB corresponds to a 5-dB OSNR [34].

The extracted 10-Gb/s payload channel is then transmitted to the 1/4-bit Mach-Zehnder DLI. The DLI used here is an off-the-shelf differential phase-shift keying (DPSK) demodulator that is commercially-available from Optoplex. The DLI supports C-band frequencies and is experimentally phase-tuned for maximum and minimum power, respectively, in the two output ports. The Optoplex DLI has a tunable FSR and is highly stable, with a specification-sheet rating of less than 1% of FSR error. The DLI then feeds a fiber switch device from Polatis with integrated power meter capabilities and an interfacing RS232 serial port. The power information (P_{const} and P_{dest}) resulting from the DLI interference is obtained by the power meters and sent using the serial interface to a high-speed Xilinx Virtex-5 FPGA. The FPGA acquires the power values every timeslot (20 ms) (once per packet) and processes the information online to calculate the packet's OSNR on a packet-by-packet basis. If the measured OSNR is greater than a predetermined threshold (here, set to 5 dB), the FPGA generates an electrical gating pulse which is transmitted to a SOA-based switch. In this experiment, the switch is controlled by a CPLD, which uses this FPGA-generated signal as an indication that the minimum performance threshold has been satisfied. If the measured OSNR is less than the threshold, the FPGA does not generate the pulse for the CPLD. Simultaneous to the OSNR measurement system, the switch

uses a fixed wavelength filter and low-speed optical receiver (similar to the photonic switching elements that comprise the fabric test-bed) to extract the QoS header bit from the wavelength-striped packet. The QoS acts as a contributing factor in triggering protection routing.

The CPLD in the SOA-based cross-layer switch contains the synthesized routing logic that makes the packet protection decision, using both the measured OSNR signal from the FPGA (denoting the packet's QoT) as well as the packet's extracted QoS class as inputs. According to these metrics, the CPLD makes the decision to either:

- electrically drive the SOA to forward low-QoS (irrespective of the OSNR) and high-QoS, low-OSNR messages to the final destination port; or
- proactively discard high-QoS, low-OSNR optical messages so that the packet protection mechanism can actuate the rerouting of these messages on a protection path.

The degraded, high-priority messages are discarded by the CPLD (*i.e.* by not gating the SOA), such that an ack can be backward propagated to the source to trigger rerouting on an alternate path. This system allows the packets emerging from the switching fabric to be dynamically monitored by the cross-layer receiving node, which the fast OSNR monitor can signal to proactively reroute the degraded packets in a high-priority data stream.

Optical packets that are passed by the cross-layer node logic at the output of the switching fabric are experimentally analyzed (Figure 5.29) using an OSA and high-speed CSA. The packet examination setup incorporates a tunable grating filter to select

one 10-Gb/s payload channel of the optical packet for system analysis and verification, which is sent to an EDFA, an optical tunable filter, and then to a VOA. The packet is then transmitted to a DC-coupled 10-Gb/s *p-i-n* photodiode and TIA, followed by a LA (RX). The electrical signals are transmitted to a BERT that is synchronized with the PPG and the fast pattern-generating FPGA; no clock recovery is performed in this experiment. The FPGA also generates an electronic gating signal that allows the BERT to be gating for BER testing over the length of the packets. The packets were gated for over 85% of their duration.

5.5.4 Results

In this experiment, a pattern of 8×10-Gb/s NRZ-OOK wavelength-striped optical packets with high and low OSNRs is injected into the test-bed (as created by the above experimental setup). The per-packet OSNR monitoring and packet protection method is shown using the Virtex-5 FPGA and SOA switch to accurately discard optical messages. The modified cross-layer receiving design measures the packets' OSNR and proactively detects (and discards) the high-priority packets with OSNR values that are specifically degraded below the performance threshold. Packets are shown to be correctly routed by the switching fabric to their desired destinations, using the DLI-based monitoring system, based on the encoded QoS level and OSNR measurement signal. The cross-layer node identifies the messages on a packet-by-packet timescale, initiating a rerouting to the degraded stream using the fabric's ack signal. The packets can then be rerouted on an alternate protection path. Several payload wavelength channels in the egressing packets are selected by the TB9 filter, exhibiting

similar results.

Using the BERT and packet-analysis system outlined previously, error-free operation of the complete system is attained, confirming that all packets at the output of the fabric and cross-layer node operate with BERs less than 10^{-12} on all eight payload channels. Thus, the system here achieves error-free operation without needing to use any FEC techniques. The power penalty performance for the experimental system is evaluated for the optical packets egressing from the three-stage switch (the worst-case path). Figure 5.30 provides the sensitivity curves for one typical payload wavelength channel ($\lambda = 1556.5$ nm). The back-to-back measurements correspond to the packet prior to injection in the fabric. The experimental system is seen to incur a 2-dB power penalty (taken at a BER of 10^{-9}), which includes the three-stage switch and the cross-layer node. The 10-Gb/s input and output optical eye diagrams corresponding to the fabric input and output are provided as insets in Figure 5.30.

In short, the fast OPM measurement capabilities here provide a means to detect and actuate dynamic rerouting of degraded, high-QoS streams. The message-level monitoring of optical packets' OSNR at the output of an OPS fabric test-bed uses a Mach-Zehnder-interferometric based approach, leveraging the dynamic measurements as a physical-layer performance indicator within a QoS-aware packet protection mechanism. The switching scheme discards degraded wavelength-striped optical messages based on the encoded priority and actuates packet rerouting on an alternate route as required. Error-free transmission of the routed high-bandwidth messages is obtained. This work explores developing cross-layer designs and routing control algorithms for future networks based on emerging real-time physical-layer measurement

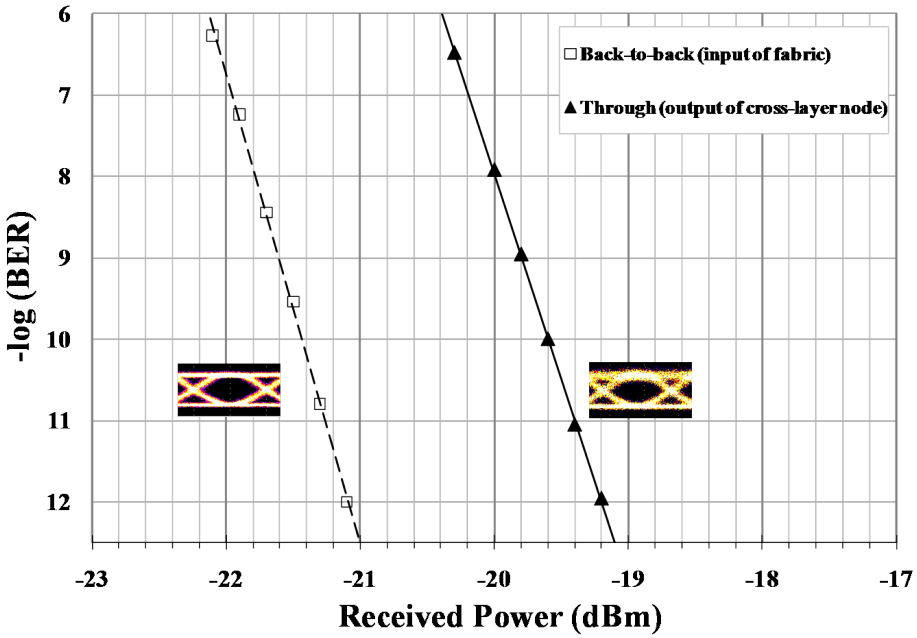

Figure 5.30: OSNR Monitoring Sensitivity Curves - Sensitivity curves at 10 Gb/s for the experiment; dashed line/unfilled points refer to the back-to-back measurements and solid line/filled points correspond to the through measurements. Insets show the 10-Gb/s optical eye diagrams (input: left; output: right).

subsystems, and incorporating varying QoS protocols. These schemes can dynamically optimize physical-layer switching, enabling a deeper exposure of the physical-layer substrate.

5.6 Real-Time Burst Sampling: TiSER

Within the overarching goal of developing fast performance monitoring modules for OPS fabrics, the work outlined in this section utilizes real-time burst sampling (RBS) [164] to provide a means of cross-layer signal monitoring in the OPS test-bed [165]. The RBS is enabled by using a photonic time-stretch enhanced recording (TiSER) oscilloscope within the test-bed, allowing for the capture of 10-Gb/s eye diagrams with a high-speed digitizer, with the future goal of rapidly extrapolating the BER for use in the cross-layer optimized infrastructure.

Leveraging eye diagrams to extract the signals' quality (Q) factor is a feasible way to realize performance monitoring. Other researchers have used captured open eye diagrams, as well as a MZ interferometer, to calculate the Q-factor of high-speed data in order to monitor the optical channels' CD [156]. Further, other similar work has been executed to estimate the BER using Q-factor monitoring [166]. Here, this work endeavors to monitor data via the eye diagram on a fast, packet-level timescale. This section deals with initial strides to this end by monitoring the 10-Gb/s payload channels that make up a wavelength-striped packet. As the next step (and as will be discussed in Chapter 6), the true high-speed capability of the RBS system will be exploited to monitor the eye diagrams at 40 Gb/s and beyond, and actuate a cross-layer feedback mechanism based on the quality of the eye.

The integrated cross-layer design will leverage emerging physical-layer technologies and systems to allow for introspective access to the optical layer. The physical layer will possess embedded real-time OPM and/or PM modules providing feedback to higher routing layers. The measurements can then result in dynamic network routing and packet protection with rapid capacity provisioning, and a means to optimize overall performance and efficiency. Here, the dedicated PM module is envisioned to monitor the BER, which can then be utilized to reconfigure optical routing using enhanced cross-layer algorithms.

OPS fabrics support the required high-bandwidth network connections for future data-centric Internet applications, by transparently supporting broadband wavelength-striped optical messages though WDM. Each wavelength channel in the multiwavelength packet will need to scale to higher data rates. As a result, the receivers in these network links will require high-speed analog-to-digital (A/D) conversion and fast digital signal processing. Thus, the bandwidth limitations of the electronic A/D converters comprise a key bottleneck in performance.

The real-time packet-level monitoring and measurement of these broadband high-speed data signals will be required for the future cross-layer-optimized platform. The photonic TiSER oscilloscope [167, 168] has been proposed as a promising technology to address this challenge by providing real-time digitization of high-speed signals, thereby realizing a true real-time diagnostic and performance monitoring tool for high-speed optical links. TiSER uses RBS to effectively slow down the signal to accommodate the digitizer's bandwidth. By embedding TiSER within an OPS fabric, the vision is a dynamic system whereby real-time eye diagrams can be generated, physical-layer

impairments can be characterized, and rapid BER measurements can be achieved.

This section discusses the implementation of the TiSER oscilloscope within the realized 4×4 cross-layer enabled switching fabric test-bed, enabling real-time performance monitoring with a message granularity. Wavelength-striped optical packets with 8×10-Gb/s payloads are correctly routed through the test-bed, and error-free performance is achieved with BERs less than 10^{-12}. TiSER captures the 10-Gb/s eye diagrams corresponding to the error-free signals at the output of the switching fabric.

5.6.1 Overview of TiSER

Designed and developed by Jalali *et al.* , the TiSER oscilloscope uses photonic time-stretch pre-processing (Figure 5.31) to perform RBS of high-speed data signals. TiSER captures a burst of samples in real-time and reconstructs the corresponding eye diagrams in equivalent-time mode. It enables the capture of fast non-repetitive dynamics at the modulation rate, comprising a real-time monitoring solution for high-data-rate optical links. TiSER has been shown to capture data signals up to 45 Gb/s [164], and more recently, up to 100-Gb/s return-to-zero (NZ) differential quaternary phase-shift keying (DQPSK) [169]. By capturing high-speed signals using commercial slower digitizers, TiSER bridges the gap in measurement functionality and performance between sampling oscilloscopes and real-time digitizers.

RBS captures bursts of measurement samples in real-time in each sampling period. Figure 5.32, from [164], compares equivalent-time sampling (using a sampling scope), real-time sampling (using a real-time digitizer), and RBS (using TiSER). It can be

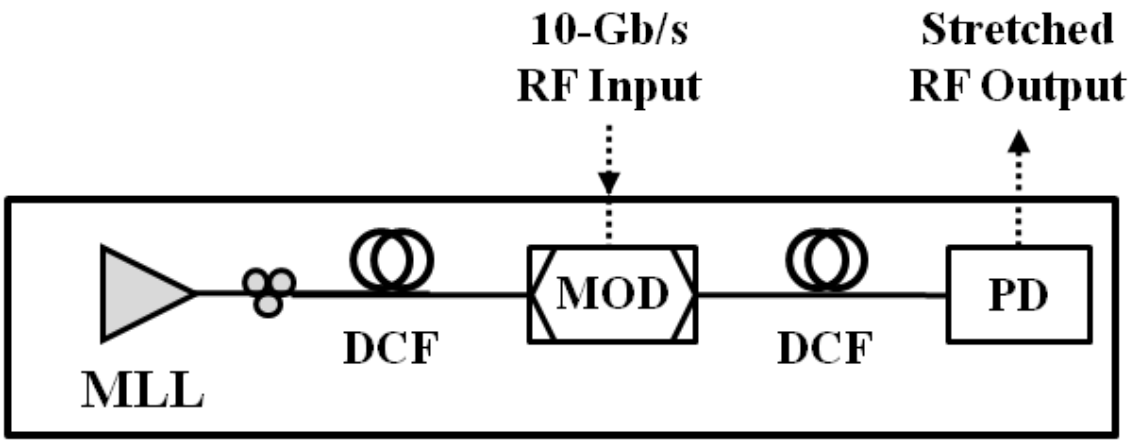

Figure 5.31: TiSER Block Diagram - Diagram illustrating the physics of the time-stretch pre-processor used by TiSER.

observed that although equivalent-time and real-time samplings are both limited in capability, RBS allows for both ultrahigh bandwidth and real-time sampling within the captured bursts.

In this experimental implementation, TiSER (Figure 5.31) uses a mode-locked laser (MLL) that generates 36-MHz ultra-short optical pulses. A -60-ps/nm dispersion-compensating fiber (DCF) then creates chirped pulses with a sufficient time aperture to support 10-Gb/s RF data rates. A Mach-Zehnder intensity modulator encodes the 10-Gb/s data signal over the chirped pulses. Propagation through a span of -657-ps/nm DCF stretches the modulated optical pulses in time, realizing a stretch factor of 12. A photodetector (PD) receives the pulses and creates an electronic RF signal that is a stretched version of the original with reduced bandwidth. A commercial A/D digitizer is used and the eye diagram is constructed using the recorded data by removing an integral number of data periods from the stretched time scale.

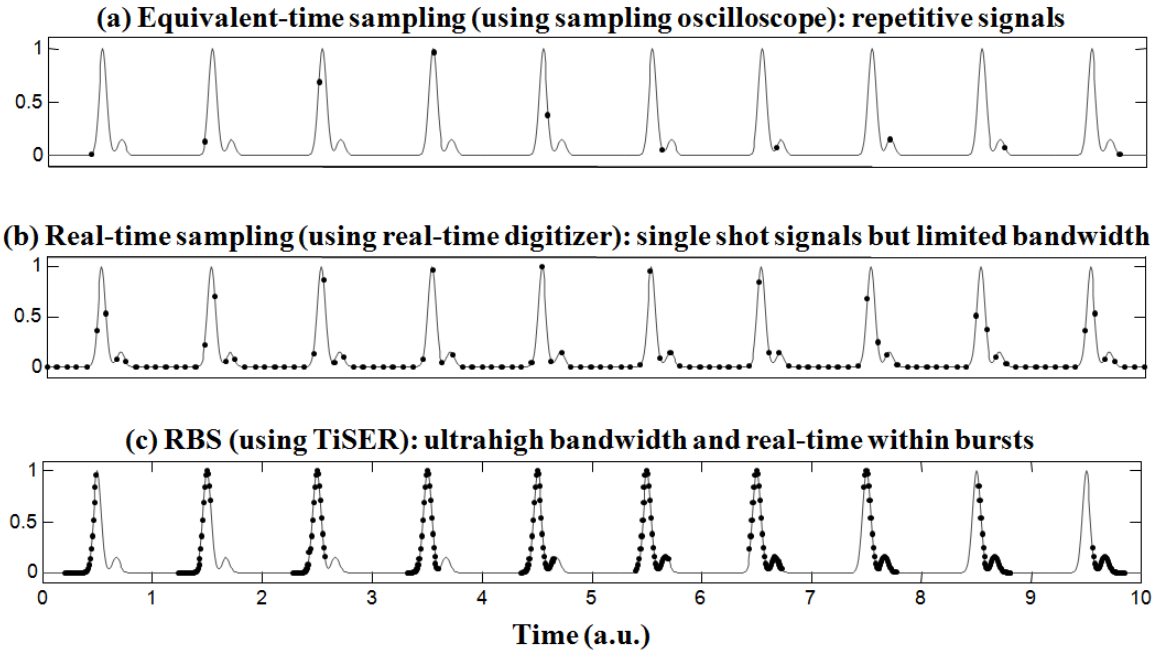

Figure 5.32: Real-Time Burst Sampling - Pictorial comparison between (a) equivalent-time sampling, (b) real-time sampling, and (c) TiSER-enabled RBS [164].

The first prototype of TiSER is implemented in a 19-inch Rackmount Chassis, which can accommodate the electronic A/D; Figure 5.33 shows a photograph of the TiSER prototype from UCLA. TiSER uses one main power supply (90 to 264 Vac, with 60-Hz universal input). All of the pre-processor components are integrated in the TiSER chassis. The inputs consist of a RF signal, a RF trigger, and a MZ modulator voltage (approximately 4 Vdc). The output ports include the stretched RF signal, the digitized data (through a universal serial bus (USB) port), and the laser clock.

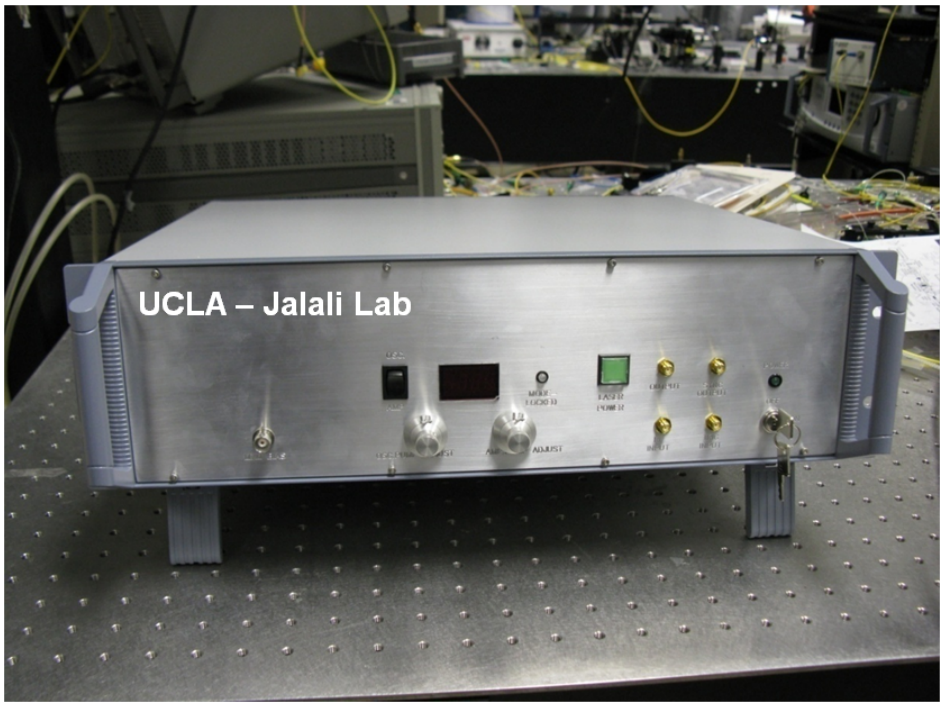

Figure 5.33: TiSER Photograph - Photograph of the initial TiSER prototype, implemented in 19-inch Rackmount Chassis.

5.6.2 Experimental Demonstration and Results

TiSER performs real-time monitoring of the optical packets propagating through a 4×4 cross-layer-enabled fabric test-bed (Figure 5.34). The fabric implementation uses six distinct PSEs that are arranged in a three-stage topology. As with previous experiments, the electronic logic within the PSEs is synthesized in the CPLDs located within each node. The system supports wavelength-striped optical packets with 10-Gb/s NRZ-OOK data on eight payload wavelength channels. Each packet also includes a four-wavelength control header. The 100-μs duration packets are modulated with a 2^{15}-1 PRBS using a single 10-Gb/s $LiNbO_3$ modulator. Here, TiSER requires a minimum of 1500 sample points per packet duration to capture the eye diagram of a single packet. Thus, the length of the optical message is actually limited by this number of samples required to sufficiently generate an eye diagram. A high-speed FPGA, which is pre-programmed with a custom test pattern, is utilized to gate the external SOAs in order to create the optical packets.

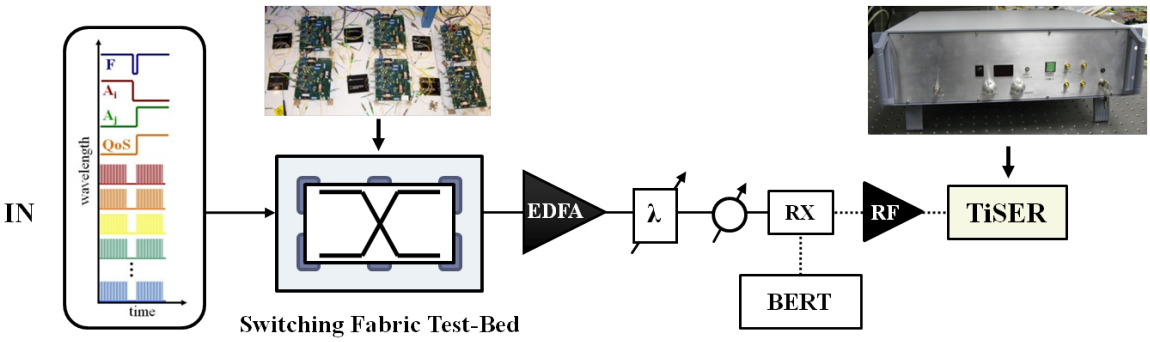

Figure 5.34: TiSER Experimental Setup - Block diagram and corresponding photographs of the experimental demonstration with the optical fabric test-bed and the TiSER box.

A pattern of optical packets is routed through the switching fabric test-bed. To evaluate the packet quality at the input and output of the test-bed, the optical packets are transmitted to a tunable optical filter, an EDFA, a second filter, a VOA, a 10-Gb/s DC-coupled *p-i-n* photodiode with a TIA and LA pair, and subsequently to the TiSER oscilloscope. Here, TiSER uses a commercially-available electronic A/D digitizer with 2-GHz bandwidth that can capture up to 20 GSamples/s. TiSER records the data samples required to generate the eye diagrams pertaining to one 10-Gb/s channel of the packet. The TiSER-captured input and output eye diagrams in Figure 5.35 correspond to one representative error-free payload channel at 10-Gb/s and are generated from a single packet (approximately 50-60 μs), illustrating the message-level granularity. It can also be observed that there is minimal degradation in the eye due to the switching fabric, as indicated by the eye diagrams in Figure 5.35.

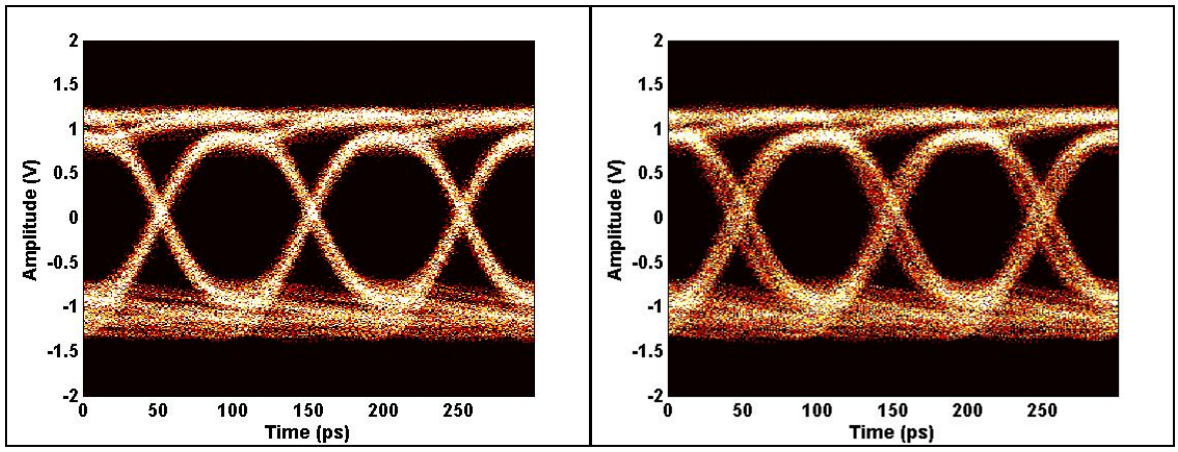

Figure 5.35: TiSER Eye Diagrams - 10-Gb/s TiSER-generated optical eye diagrams pertaining to the packets at the network input (left) and output (right) ($\lambda = 1556.6$ nm).

The accurate routing of the 8×10-Gb/s wavelength-striped optical packets through the switching fabric test-bed is verified. At the output, the electronic data received

from the 10-Gb/s receiver is simultaneously transmitted to a BERT. The BERT is gated to analyze the 10-Gb/s data over the packet durations using an electronic gate signal from the FPGA. All received packets are confirmed error-free on all eight payload wavelengths, with achieved BERs less than 10^{12}. Figure 5.36 depicts the sensitivity curves of one error-free channel, showing a power penalty less than 1 dB.

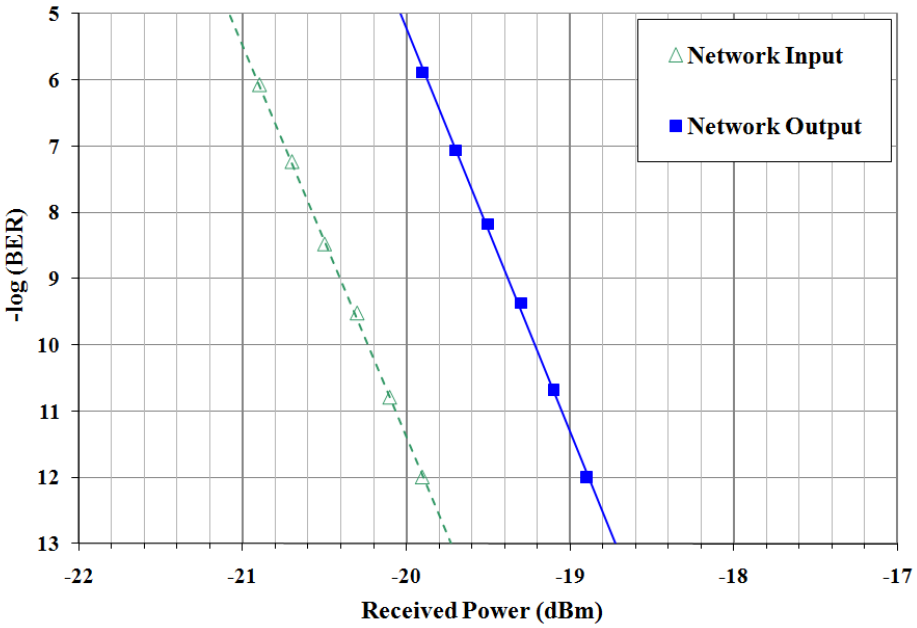

Figure 5.36: TiSER Sensitivity Curves - 10-Gb/s BER sensitivity curves for the integrated TiSER and OPS fabric operation (λ = 1556.6 nm).

In closing, the TiSER oscilloscope provides a feasible means to realize the real-time message-granular monitoring of broadband data and to enable dynamic packet routing capabilities that will be required in future cross-layer networks. This work demonstrates TiSER-generated 10-Gb/s eye diagrams of error-free 8×10-Gb/s optical packets propagating through an implemented OPS fabric test-bed.

The system shows the potential of rapidly and dynamically extrapolating the

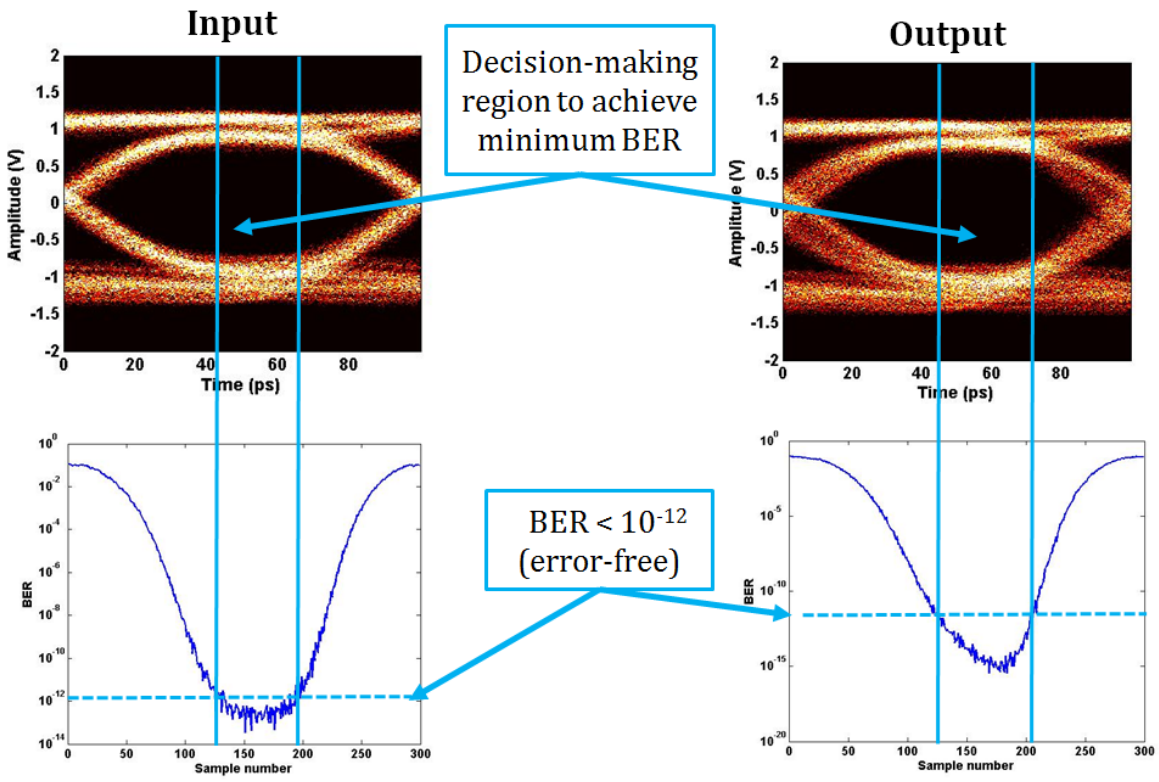

Figure 5.37: TiSER BER Extrapolation - Preliminary results from offline BER estimation algorithm.

messages' BER, which can then be used an indication of the physical-layer performance. The BER estimation will be performed using advanced signal processing, to rapidly determine the Q factor from the captured eye diagrams. Preliminary results from an offline BER estimation algorithm are shown in Figure 5.37. The region of the largest eye opening is used to measure the lowest (best) BER value.

The real-time extrapolation via online signal processing using a FPGA comprises some significant ongoing work. However, once completed, this will allow the BER of optical packets to be evaluated much faster than using a conventional BERT system, in addition to be measured on a packet-by-packet basis. The vision is to use the measurements as a characterization of real-time physical-layer performance in a cross-layer-optimized platform and the corresponding routing algorithms.

5.7 Failure Recovery

One key functionality that must be demonstrated for future optical switching fabrics is the ability to perform a fast switching and path provisioning in the face of failure. This is particularly important for the "CIAN Box" (the CLB), which will need to recover and potentially route around signal impairments. The following section of this thesis describes a demonstration of cross-layer failure recovery leveraging a fast, reconfigurable optical switching fabric with a nanosecond-scale recovery response time. Experimentally, the failure recovery capability is given by virtue of an implemented control and management plane, which is realized here using a FPGA, in conjunction with the high-switching-speed OPS fabric test-bed.

A clear advantage of OPS is its seamless ability to realize dynamic optical-layer

switching functionalities while simultaneously supporting high-throughput traffic in an energy-efficient manner [85]. OPS fabrics can be deployed in future routers to enable the transparent switching of multiwavelength optical messages.

Within the scope of the bidirectional cross-layer optimized infrastructure, the transport and routing of optical packets is affected by real-time performance monitoring metrics, which can be extracted on a packet-level timescale [30] (and discussed above). The proposed cross-layer signaling platform will allow for a unique means of dynamically optimizing network performance and energy consumption. In addition to taking into account the introspective cross-layer awareness of the physical layer, it is also necessary for the switching state of the OPS fabric to be affected by higher-layer IP parameters in a programmable fashion. The major cause of complete network failure is at the IP layer (*i.e.* when an IP router fails). Further, it has been proposed that a significant reduction in power consumption can be achieved by allowing for the routers to turn off (or be set in sleep mode) [79]. Thus, in the case of these costly router failures and/or when a router is sleep mode for energy-saving benefits, connections and lightpaths between end nodes can be maintained by optically routing around the failures and/or the idle routers. Optical-layer recovery can also realize optical restoration mechanisms that protect clients against IP router failures [170]. To enable an optical bypass [77] and prevent network outage, an on-the-fly reconfiguration of the physical-layer switching fabric will be required to mitigate failure.

In short, the optical switching fabric should be capable of executing a fast, nanosecond-scale reconfiguration of its switching state, allowing for the dynamic management of optical packets and the seamless recovery of the fabric, in the case of

IP-layer router failures and cross-layer enabled optical-layer signal degradations [171].

This work demonstrates a reconfigurable optical switching fabric architecture that dynamically responds to failure by performing a seamless recovery based on external input signals (Figure 5.38). Upon the detection of a higher-layer router failure or degraded optical packet streams, the switching fabric executes a fast, nanosecond-scale reconfiguration of its switching state to yield highly dynamic management of the optical packet routing. The cross-layer failure recovery scheme is experimentally implemented on a 4×4 multistage OPS fabric test-bed. An external FPGA device acts as an optical control and management plane for the OPS fabric, allowing the signals from either higher network layers or embedded physical-layer OPM devices to actuate failure recovery and rerouting of messages. The successful routing of 10×10-Gb/s wavelength-striped optical packets is demonstrated, for both cases of an online router (*i.e.* packets are transmitted correctly) and an offline router (*i.e.* the router or subsequent optical link is down, thus packets are rerouted according to the recovery switching logic). Based on the state of an upper-layer router, the switching fabric supports the correct routing and error-free transmission of the multiwavelength optical packets, with BERs less than 10^{-12} on all ten payload channels, with less than 1 dB of power penalty.

5.7.1 Failure Recovery Scheme

The failure recovery experiment here is performed on a 4×4 synchronous, multi-terabit capacity, multistage OPS fabric test-bed (Figure 5.38) [26]. The two-stage design uses four non-blocking 2×2 PSEs. The payload is encoded on ten payload channels, which

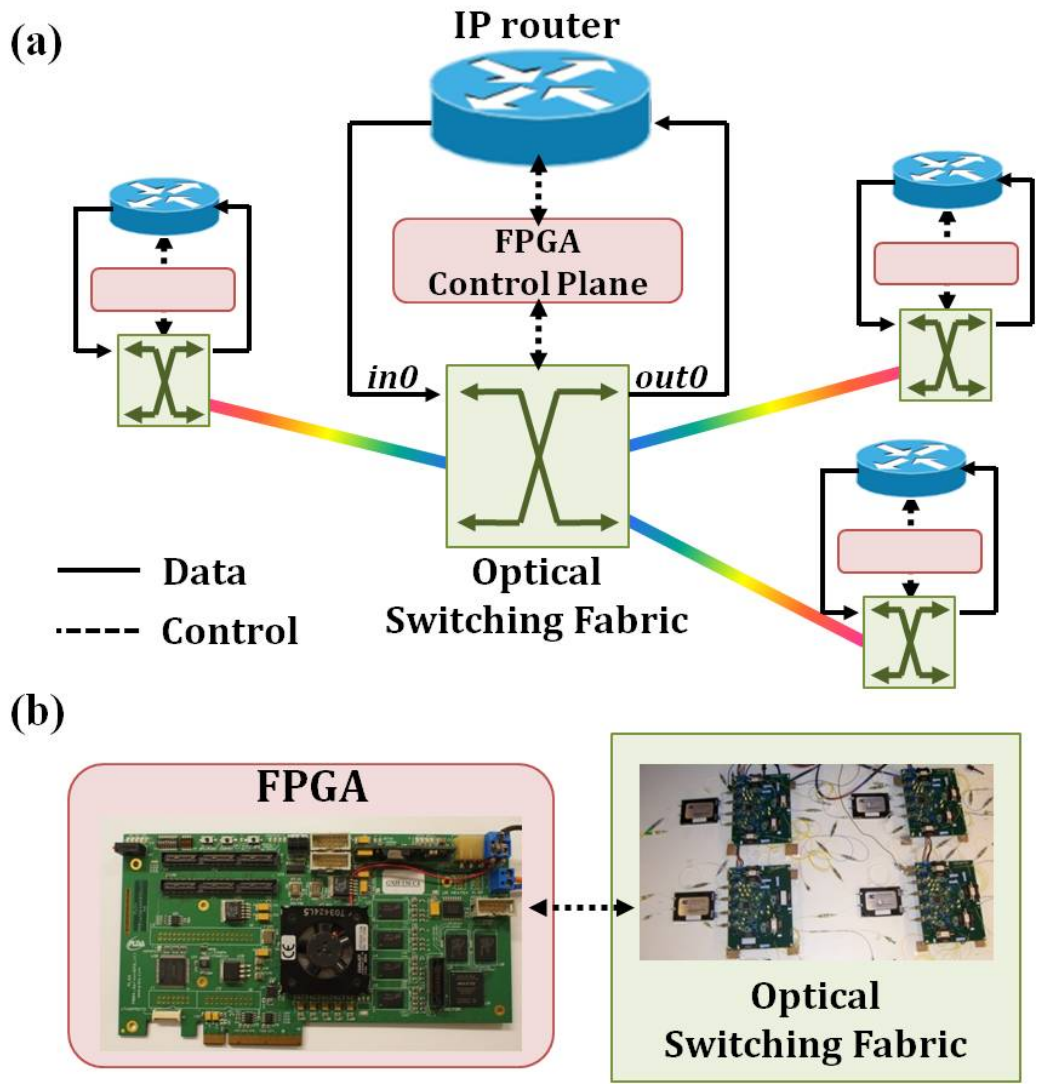

Figure 5.38: Failure Recovery Network Architecture - (a) Envisioned network architecture with network nodes composed of IP routers, optical packet switching fabrics, and a FPGA-based control and management plane. The FPGA acts as a cross-layer interface, accepting control inputs to manage physical-layer switching; (b) Photographs of FPGA circuit board and implemented OPS fabric used in this demonstration.

are modulated at 10 Gb/s (per wavelength). The 10×10-Gb/s optical packets have an aggregate bandwidth of 100 Gb/s, with each payload channel being 120 ns in length, resulting in 1500-byte messages, analogous to the Ethernet MTU.

The failure recovery scheme allows the 2×2 PSEs within the OPS fabric to be aware of – and account for – the failure of the higher-layer router, or degrading impairments on the optical-layer lightpath. Upon the detection of a failed/degraded link, the fabric recovers by reconfiguring its switching state (on a nanosecond timescale) to route around the point of failure, yielding more dynamic switching. A Stratix II GX FPGA realizes a control plane for the optical fabric that accepts external inputs (*e.g.* electronic signals from a router) and then generates failure signals for the switching nodes. The routing logic synthesized within the CPLDs is adapted to accept these electronic failure signals to either route normally (if the router is online, optical packets should be switched accordingly), or route around the failure (if the router is offline or failed, packets are rerouted to ensure that no messages are to be transmitted to the router). As in Figure 5.38, if the router is offline, packets that would have been transmitted to the router are instead rerouted to another output port if there is no contention; otherwise, they are dropped.

5.7.2 Experimental Demonstration and Results

The experiment uses the FPGA-based control plane, as well as an implemented 4×4 OPS fabric that is built with commercially-available parts. 10×10-Gb/s wavelength-striped optical packets are injected in the fabric (Figure 5.39a) and routed based on the router failure state as denoted by the control plane. A FPGA-based circuit board

is utilized, containing eight flip switches and 28 general purpose input/output (GPIO) pins. In this demonstration, the FPGA is programmed to receive input from the flip switches, indicating the presence of a router failure, and to signal the appropriate switching nodes using the GPIO pins.

The optical packets are generated using 13 separate CW-DFB lasers: three DFBs are used for the control headers (frame: 1555.75 nm; addresses: 1531.12 nm and 1543.73 nm), and ten DFBs, ranging from 1533.47 nm to 1564.68 nm, create the payload wavelength channels. A single $LiNbO_3$ modulator encodes a 2^7-1 NRZ-OOK PRBS on each of the payload channels. The CW control and the PRBS payload wavelengths are multiplexed to create a multiwavelength signal, which is then sent to a 1:3 passive splitter to form three independent input streams. Each stream is separately gated into a pattern of 1500-byte packets using external SOAs (one stream for each fabric input port) (Figure 5.39a).

Figure 5.39 provides the input and output optical waveforms pertaining to the experimental traffic sequence, denoting the frame, address, and one 10-Gb/s payload channel for all input (Figure 5.39a) and output (Figure 5.39b/c) ports. We show two explicit cases for which all packets are shown to be correctly routed to their desired outputs. First, Figure 5.39b gives the output traces for an online router scenario, where all output ports are available and packets are transmitted correctly. Second, an offline router scenario is assumed (Figure 5.39c), in which the output port corresponding to the router is designated as failed (or possibly in sleep mode). The FPGA informs the OPS fabric of the failure, thus the fabric can reconfigure its switching state with a nanosecond response time to mitigate the failure and reroute traffic to avoid sending

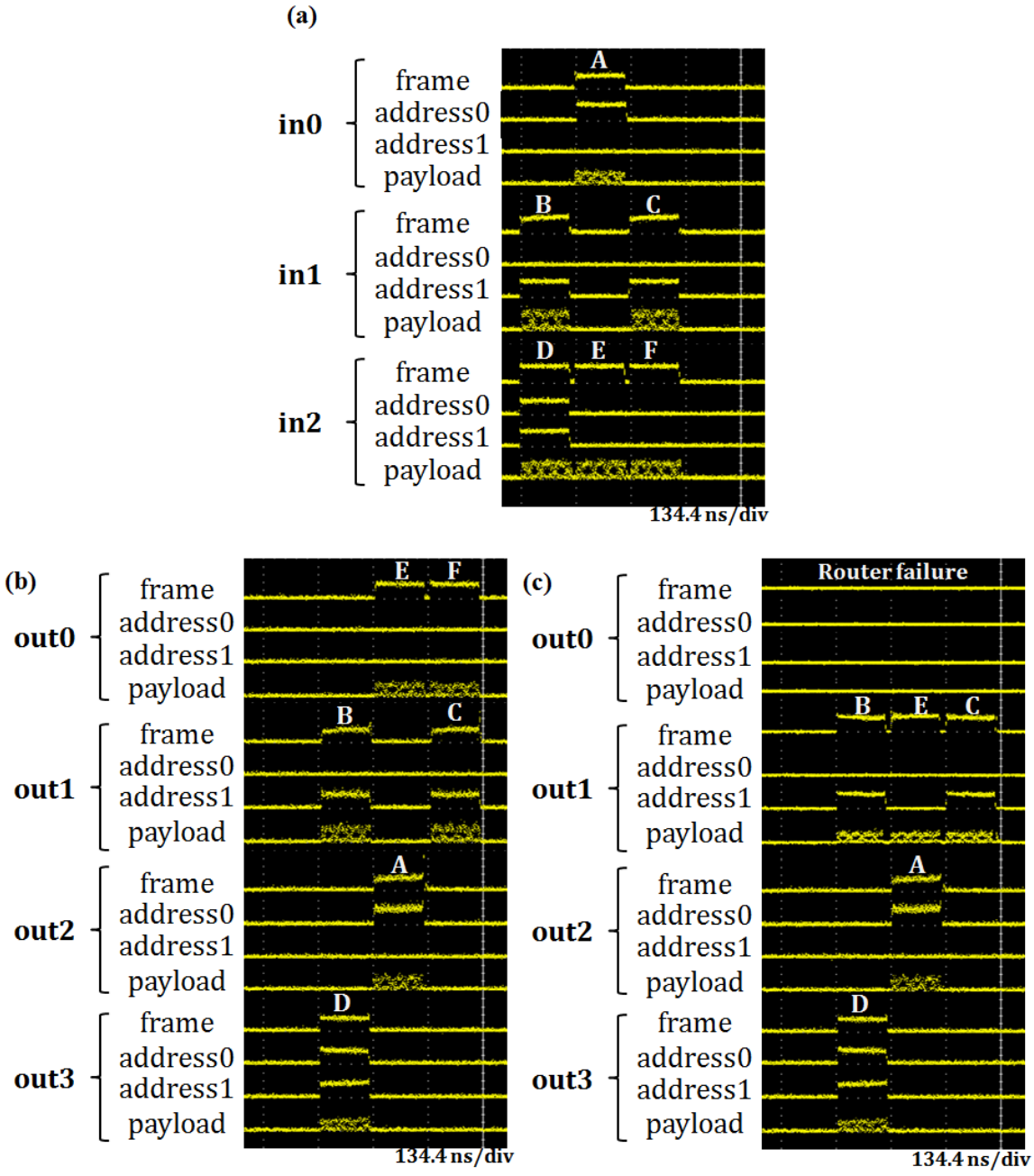

Figure 5.39: Failure Recovery Waveform Traces - Optical waveform traffic traces of the experimental packets taken at the (a) fabric input; (b) fabric output under an online router scenario with all output ports functional; (c) fabric output in the case of an offline router, with switching fabric aware that link out0 is down/degraded.

packets to the failed node. In Figure 5.39, it is seen that no packets are transmitted to out0 (to the router); messages formerly intended for out0 (*i.e.* packets E and F) are rerouted to out1 (if the port is available). Here, packets C and F contend for out1, thus F is dropped. The logic in the PSE prioritizes messages originally designated for the next node and chooses to drop messages that were originally being forwarded to the port associated with the router.

All 10×10-Gb/s wavelength-striped packets are shown to be transmitted error-free at the fabric's output. Signal integrity is verified using a DC-coupled 10-Gb/s *p-i-n* photodiode and TIA followed by a LA, and a BERT that is synchronized with the packet gating signals. The BERT is gated to provide error checking on the length of the optical packets. BERs less than 10^{-12} are attained on all ten payload channels. Using a high-speed CSA, the optical eye diagrams are examined for all ten data payloads. Figure 5.40 shows the 10-Gb/s input and output eye diagrams for all ten of payload wavelength channels. The BER sensitivity curves for $\lambda = 1564.68$ nm, the payload channel exhibiting the most degraded optical eye, is given in Figure 5.41. The two-stage fabric shows a 0.9-dB power penalty (yielding a 0.45 dB power penalty per SOA hop), taken at a BER of 10^{-9}.

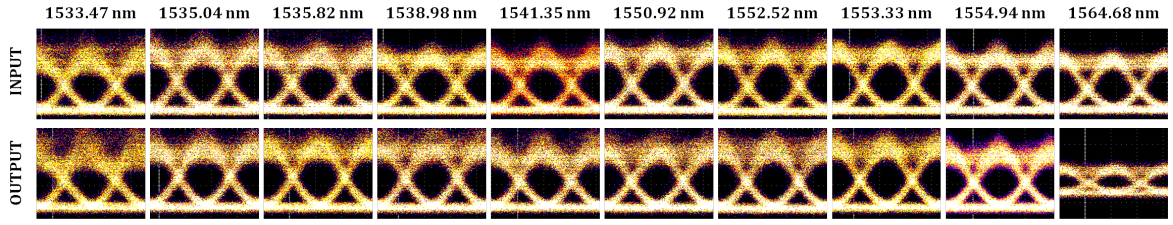

Figure 5.40: Failure Recovery Eye Diagrams - 10-Gb/s optical eye diagrams for the input and output of the switching fabric, for all ten payload wavelength channels.

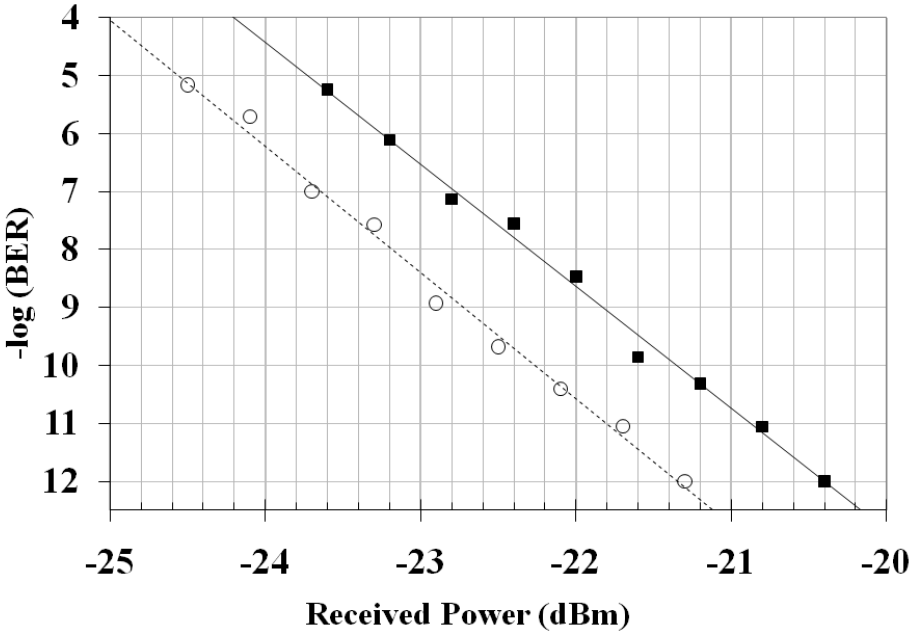

Figure 5.41: Failure Recovery Sensitivity Curves - BER sensitivity measurements for one payload channel ($\lambda = 1564.68$ nm). The dashed line with open points corresponds to the back-to-back measurements, while the solid line with filled points refers to the data at the fabric's output.

In closing, future deployed OPS fabrics will likely need to exhibit an advanced level of agility and dynamics in the face of both higher-layer failures and optical signal degradation. This work demonstrates an OPS architecture that can seamlessly perform a fast, nanosecond-scale recovery from router failures by leveraging a cross-layer control plane to realize enhanced optical-layer switching functionalities. The switching fabric is reconfigured on-the-fly using FPGA control signals, providing a means of protecting packet transmission and realizing optical bypasses that allow traffic to bypass failed nodes. The fabric supports the correct, error-free transmission of all wavelength-striped optical packets. This demonstration comprises a fundamental step to a fully operational cross-layer optimized CIAN CLB node [20] (and as discussed previously in Chapter 2) that features a reconfigurable OPS fabric with dynamic response to failures, a FPGA-based control and management plane, and emerging performance monitoring devices, to provide innovative, low-cost technologies for future access/aggregation networks.

5.8 Closing Remarks

The ultimate goal of the work discussed in this chapter is to enable a more intelligent, dynamic, and programmable optical layer using the proposed cross-layer approach. Emerging optical technologies (including optical packet switching and performance monitoring subsystems) are used to create a bidirectional cross-layer communication design that can optimize fabric operation incorporating introspective knowledge of the optical flows' quality. This platform evaluates optical signal degradations and leverages packets' quality-of-transmission in order to provide a real-time feedback to higher routing layers. Allowing the physical layer to interact dynamically with upper

layers will enable enhanced routing for critical data flows.

This chapter illustrates the author's work in designing and developing a cross-layer infrastructure that can leverage physical-layer introspection. The platform is first demonstrated with a message control interface, managing the injection and transmission of packets within the switching fabric. Subsequent work in developing the cross-layer communications supported infrastructure features a custom receiving node on the implemented OPS fabric test-bed. As an example of cross-layer signaling, proactive packet protection mechanisms and QoS-based multicasting are investigated, both in an experimental environment and in simulation. The complete system will require dedicated performance monitoring modules embedded in the optical layer to assess the physical-layer performance. Thus, a few demonstrations of the fast measurement techniques are further explored here, in collaboration with several CIAN-related research groups. An optical control plane to support the routing algorithms and to facilitate fast physical-layer recovery is also demonstrated.

All of this work described in Chapter 4 and Chapter 5 cumulates in an integrated experimental demonstration that will allow the cross-layering schemes to actuate packet-level or flow-based rerouting using a small-scale network of CIAN CLBs with fast OPM modules. An initial prototype of the CLB is covered in the next chapter (Chapter 6), discussing the first demonstration of a CIAN Box. The fast recovery of optical-layer messages will be performed upon the detection of degraded optical signals. The CLBs will utilize the optical switching fabric, dynamic monitoring functionalities, and a FPGA-based control plane, and will showcase the transmission of "real" data traffic via the support of a streaming video application. The demonstration of a mesh-

centered network composed of multiple CIAN CLBs is ongoing at the time of writing of this dissertation.

Chapter 6

Cross-Layer Network Node

The following chapter presents the ultimate capstone of this dissertation's work, consisting of the development and realization of the first prototype of the cross-layer enabled network node: the CIAN cross-layer box (*i.e.* the CIAN Box or CLB). The cross-layer box is first proposed by the author in this dissertation, and an initial prototype is discussed herein.

The design of the cross-layer network element node builds upon all the previously presented work on utilizing the optical switching fabric test-bed, supporting advanced network functionalities, and developing a unique environment for cross-layer communications and optimization. This demonstration embodies the pinnacle of the author's work, acting as a point of culmination of various research activities described previously in Chapters 3, 4, and 5.

6.1 Goals

As the data rates of the optical channels increase to meet the high-bandwidth demands, all-optical transmission and switching may become the technology of choice for future networks. However, while this may allow the infrastructure to reach ultrahigh capacities, by reducing the number of O/E/O conversions and utilizing a reduced percentage of electronic components, the system will lose access to electronic regeneration and grooming functionalities. Since these are important to preserve adequate signal integrity for end-to-end network links, this transformation results in the overall network becoming much more sensitive to physical-layer impairments.

The ultimate goal is to enable the real-time monitoring of the health of the optical channels, as well as realize on-the-fly reconfiguration and recovery on a packet-timescale. By leveraging packet-level monitoring using modules embedded in the physical layer, the CLB will provide advanced awareness of the optical properties and channels to ensure the data signals maintain a high QoT in the optical domain. In this way, the cross-layer box allows the network to support networking and routing decisions for the optical data, which is an innovative characteristic of the CIAN Box as compared to current aggregation nodes.

The CLB will be able to support heterogeneous aggregation traffic and high-bandwidth applications, with varying levels of QoS, optimizing the performance of the switched optical data. The option to react to the awareness of the optical channel properties and performance at a packet-rate timescale can also be dependent on energy- and QoS-aware algorithmic inputs. In the future, the final cross-layer platform will realize various dynamic routing applications and support various multi-

layer optimization and traffic engineering protocols, in order to allow for the co-optimization of QoS and QoT with energy awareness [172, 173].

In short, the current vision for the CIAN cross-layer box as a novel intelligent optical aggregation network node features the following components:

- packet-rate reconfiguration and optical switching capabilities;
- advanced physical-layer functionalities;
- dynamic performance measurement subsystems;
- a distributed cross-layer control plane; and
- cross-layer network routing protocols enabling dynamic resource allocation and multi-layer traffic engineering.

Figure 6.1 shows a block schematic of the cross-layer enabled node (a more detailed view than in Figure 2.4). The development of an integrated CLB node will result in the cross-layering techniques being able to actuate packet-level or flow-based rerouting, as well as the support for high-bandwidth applications (*e.g.* HD video streaming). Figure 6.1 depicts the current view of the various components that make up the box; the specific functions of the node will undoubtedly evolve to handle the critical needs of industry, the networking community, and CIAN.

6.2 First Prototype

An initial prototype of the CIAN cross-layer box is described here for the first time. Figure 6.2 provides an overview of the version of the cross-layer box implemented in the

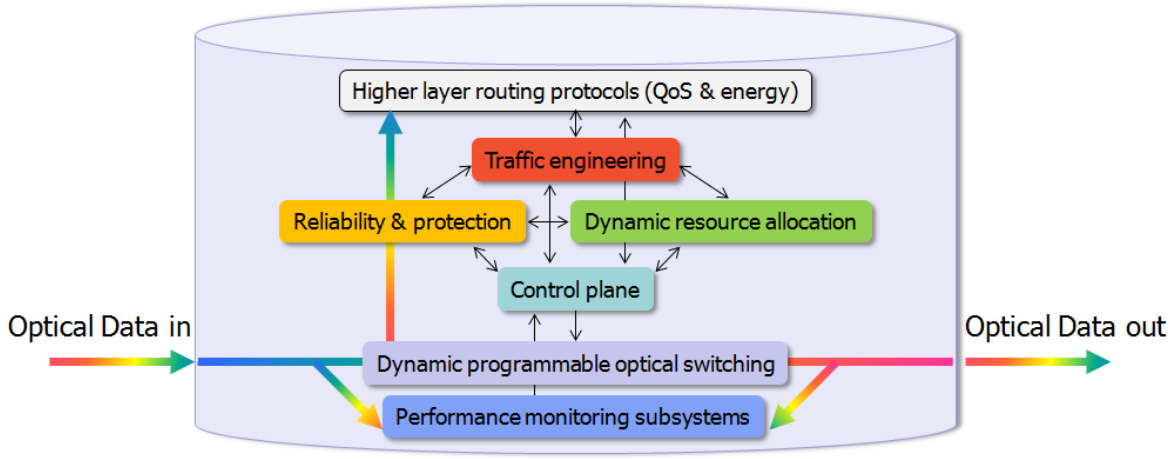

Figure 6.1: Detailed Cross-Layer Box - Detailed view of the cross-layer box with various components.

following experimental demonstration. The three major components that have been realized thus far are depicted.

In its current test-bed implementation, the cross-layer box is composed of the following components:

- a dynamic programmable optical switching fabric to enable fast all-optical switching of wavelength-striped optical messages (purple block in Figure 6.2);
- packet-level performance monitors to evaluate the optical data (dark blue block in Figure 6.2); and
- a FPGA-based control plane to support packet-rate reconfiguration and feedback from the optical layer (light blue block in Figure 6.2).

Utilizing these components, the current capabilities of the CIAN Box include the detailed measurements of the optical packets' BER and OSNR in order to enable

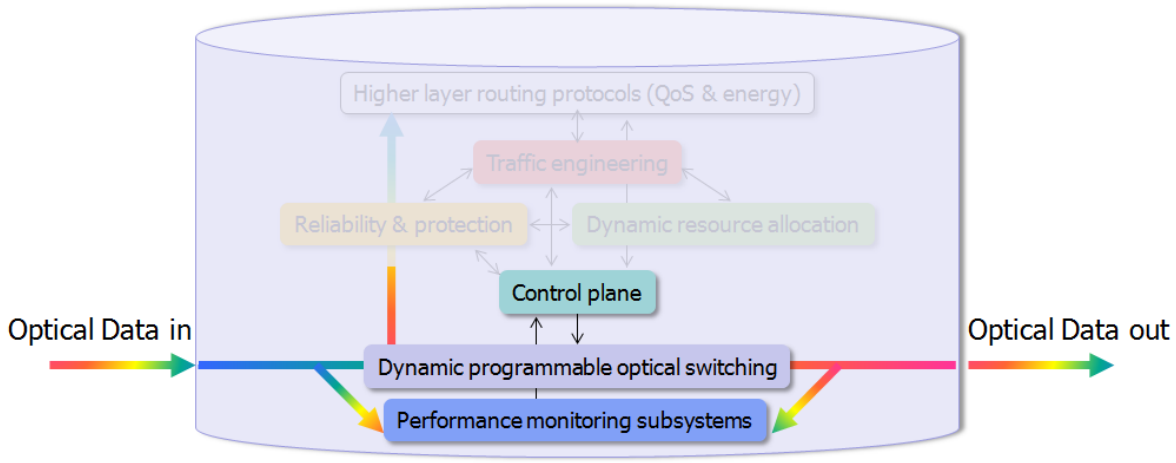

Figure 6.2: Detailed Cross-Layer Box with Demonstrated Capabilities - Detailed block schematic of the implemented cross-layer box, showing the components that have been realized in the experimental demonstration. The faded blocks have not yet been implemented in the box and constitute future work.

proactive packet protection switching. Additional adaptations include optical packet multicasting and other advanced switching functionalities (as discussed in previous chapters).

Here, a few select functionalities have been chosen to showcase the capabilities of this first CLB prototype. The switching fabric within the CLB is shown to support the aggregation of multiple data rates through the simultaneous transmission of:

- 8×40-Gb/s wavelength-striped optical packets with NRZ-OOK PRBS, and
- 4×3.125-Gb/s 10-Gigabit Ethernet-based (10GE-based) HD video data.

This particular dimension of the experiment also shows the support for concurrent packet- and circuit-switched lightpaths within the switching fabric at a given time (similar to [174]).

The fast packet-scale reconfiguration of the switching fabric is demonstrated in this experiment using the FPGA-based control plane in two distinct cases, *i.e.* the fabric is shown to reconfigure quickly while supporting two data streams with different data rates. First, the QoT of high-bandwidth optical packets (using one of the supported 40-Gb/s optical payload channels) is shown to be assessed using TiSER [168] at one of the output ports. Upon the detection of a failure or a degraded link (*i.e.* as indicated by the TiSER-extrapolated BER), the control plane then signals the switching fabric to switch paths to reroute the optical packets, in order to dynamically avoid the impairment.

Second, a 10-Gigabit Ethernet-based optical network interface card (O-NIC) is leveraged to allow for the transmission of circuit-switched video data through the switching fabric without distortion or frame loss. Again, in the face of higher-layer router failure and/or the detection of optical signal degradation, the FPGA control plane can then allow the switching fabric to perform a nanosecond-scale recovery and allow the video data to be transmitted seamlessly upon restoration of the optical link. Further, the cross-layer adaptability of the application layer to the physical layer is demonstrated using variable-bit-rate (VBR) video transmission over the switching fabric.

6.3 Implementation Overview

Figure 6.3 depicts the complete envisioned network architecture, deploying several CLBs in a mesh-type topology. The bidirectional signaling infrastructure interconnecting the boxes with the FPGA-based control plane and potential edge users is shown. The control plane is used to interface the cross-layer boxes and

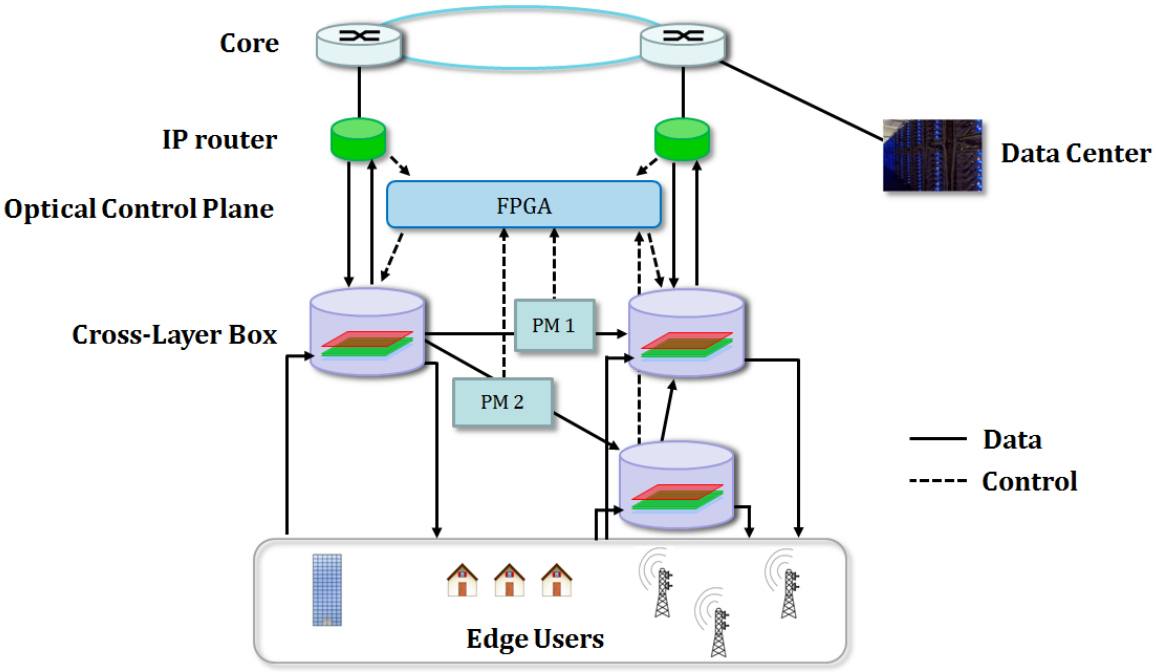

Figure 6.3: Architecture of CLB-Enabled Network - Envisioned cross-layer box enabled network architecture, depicting the bidirectional signaling infrastructure resulting from the cross-layer interactions. The FPGA represents the control plane and PM blocks denote the performance monitoring subsystems.

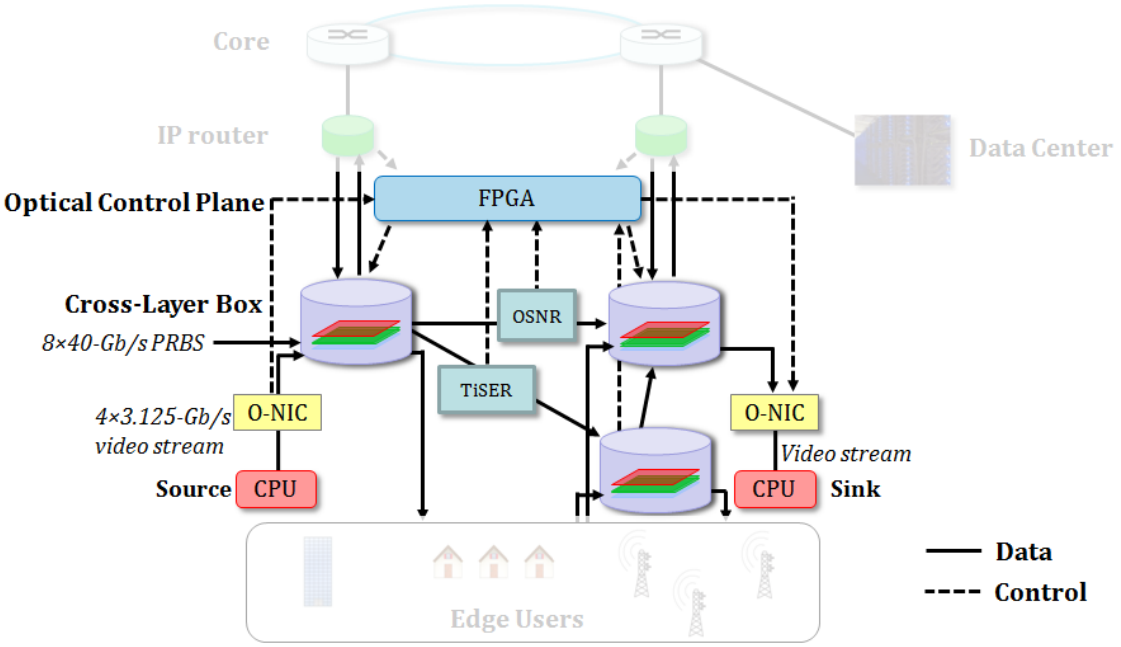

Figure 6.4: Demonstrated Architecture of CLB-Enabled Network - High-level overview of implemented cross-layer box enabled network. Demonstrated capabilities including the support for high-bandwidth 8×40-Gb/s optical packets, and 10GE-based video streams. Faded blocks will be realized in subsequent work.

physical-layer performance monitoring modules with the higher-layer router nodes. Figure 6.4 subsequently shows the focus and implemented parts of this experimental demonstration, with faded network connections and entities for components that were not realized here.

This work incorporates multiple previously-presented functionalities of the switching fabric and aspects of the cross-layer platform (in addition to several novel realizations) to exemplify and achieve a single experimental demonstration of a CIAN Box. The implemented features of the current CLB include:

- a multi-terabit capacity switching fabric,
- fast reconfiguration and recovery from failures,
- a cross-layer control plane (*i.e.* via the FPGA), and
- a performance monitoring subsystem (*i.e.* the TiSER oscilloscope),

while supporting:

- the aggregation of multiple data rates, and
- the VBR transmission of 10GE-based video.

These functionalities operate on the timescale of a single optical packet (*i.e.* ∼10s of nanoseconds) to minimize traffic loss and packet dropping, as compared to ∼10s of milliseconds for current aggregation systems. Other capabilities such as QoS-based switching, packet multicasting, monitoring of other optical channel properties, *etc.* may also be demonstrated in future work.

6.4 Experimental Demonstration

The demonstration of the initial prototype of a CIAN Box is outlined in the following section of this thesis. This first demonstration of video transmission on a multi-terabit cross-layer enabled network node features a reconfigurable optical switching fabric and packet-level performance monitoring. The prototype is shown to support and aggregate the data from a high-bandwidth source (*i.e.* the 8×40-Gb/s wavelength-striped packets), with the transmission of video using a 10GE O-NIC (*i.e.* the 4×3.125-Gb/s video data).

The central theme for the experiment is the fast nanosecond-scale recovery and reconfiguration of the box's switching fabric upon the detection of either a failed higher-layer router and/or degraded optical signals. In this way, the optical-layer data can be rerouted within the switching fabric to maintain a high QoT as determined by the embedded performance monitor (in this case, TiSER).

This per-packet reconfiguration of the switching fabric uses the FPGA-based control plane in a two-part experiment, wherein both parts are occurring concurrently. The optical fabric is operated simultaneously with two aggregated streams of traffic at different data rates, and is shown to recover and reconfigure at the packet rate (∼10s of nanoseconds). The first part of the demonstration leverages the large multi-terabit capacity of the switching fabric, as well as the ability to implement the TiSER oscilloscope to perform the monitoring of a single 40-Gb/s payload channel, while the second utilizes a 10GE-based O-NIC to support the transmission of HD video. The following sections discuss the experimental demonstration in detail, providing the experimental setups and results of these corresponding parts.

6.5 Fabric Reconfiguration

The fast reconfiguration capability of the optical switching fabric within the cross-layer box has been discussed above (and also in [171]). Figure 5.38 depicts the architecture. Here, a similar experimental setup is used.

Upon the occurrence of a higher-layer router failure or the detection of a degraded optical link, the switching fabric executes a fast, packet-rate reconfiguration of its switching state. The reconfiguration experiment is performed on a 4×4 multi-terabit capacity, multistage OPS fabric test-bed [26]. The two-stage implementation utilizes four non-blocking 2×2 PSEs that are constructed from commercial off-the-shelf components (described in detail in Chapter 3).

A separate FPGA device realizes the cross-layer control, allowing the bidirectional signaling from the higher network layers to affect optical-layer switching. The control plane leverages a commercial FPGA circuit board, with eight flip switches and 28 GPIO pins. In the experiment, the FPGA is programmed to receive input from the flip switches, indicating the presence of a router failure, and to signal the appropriate PSEs using the GPIO pins. The photograph in Figure 6.5 depicts the test-bed environment with the implemented 4×4 optical switching fabric and the FPGA-based control plane.

Here, the operation of the CLB's switching fabric supports the simultaneous transmission of 8×40-Gb/s wavelength-striped optical packets and the 4×3.125-Gb/s multiwavelength HD circuit-switched video streams, with the FPGA control plane signaling the need for fast recovery. Similar to past packet-scale reconfiguration experiments, two cases are shown: an online router (with packets being transmitted correctly), and an offline router (with packets being rerouted according to the recovery

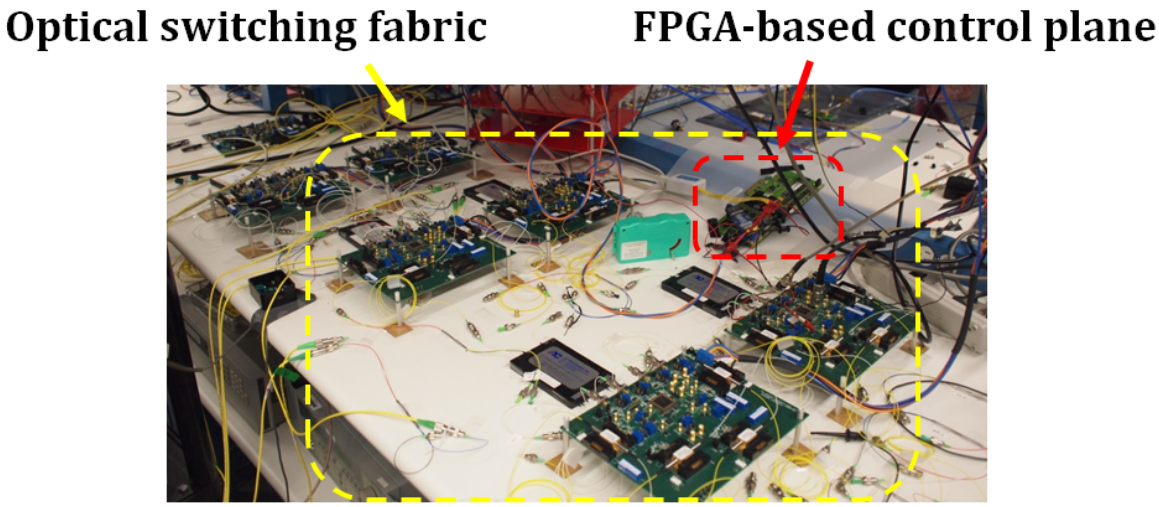

Figure 6.5: Cross-Layer Box Photograph - Photograph showing the implemented 4×4 optical switching fabric and FPGA-based control plane comprising the CLB.

switching logic upon the detection of a failed router or degraded subsequent optical link).

A detailed experimental setup for the complete demonstration is shown in Figure 6.6, with the setup corresponding to the fabric's recovery of 8×40-Gb/s data using TiSER (shown in blue) and the setup corresponding to the failure recovery of the 10GE video streams (shown in green).

6.6 Multi-Terabit Fabric Reconfiguration with TiSER

The following section describes the fast reconfiguration of the switching fabric as it operates with a potentially multi-terabit load. The switching fabric supports 8×40-Gb/s wavelength-striped optical packets [26], which are injected in the fabric and

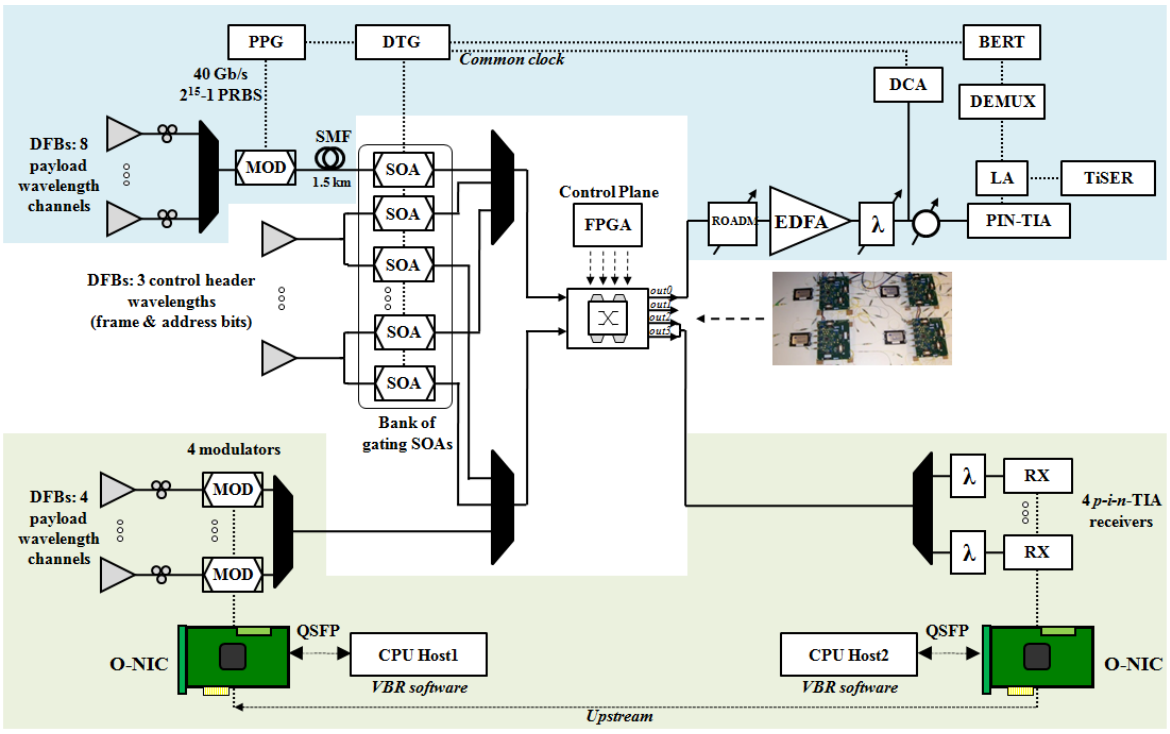

Figure 6.6: CLB Demonstration Experimental Setup - Detailed experimental setup diagram of the complete demonstration. The blue region denotes the setup associated with the 8×40-Gb/s packet generation; the green region denotes the setup associated with the 4×3.125-Gb/s video data.

switched depending on the router failure state as signaled by the FPGA-based control plane.

6.6.1 Experimental Setup

The payload information here is encoded on eight separate payload channels, which are each modulated at 40 Gb/s (per wavelength channel). The 8×40-Gb/s optical packets have a total aggregate bandwidth of 320 Gb/s (per fabric input port), showcasing the multi-terabit capacity of the switching fabric.

The blue region in Figure 6.6 depicts the setup for the 8×40-Gb/s packet generation and signal integrity analysis. The front- and back-end systems are similar to that of other previous experiments. The payload channels are generated using eight separate CW-DFB lasers; the following wavelengths are used: 1533.12 nm (C58), 1535.04 nm (C53), 1538.98 nm (C48), 1541.35 nm (C45), 1550.92 nm (C33), 1552.52 nm (C31), 1553.33 nm (C30), and 1560.61 nm (C21). The minimum frequency spacing between two adjacent payload channels is 100 GHz (or 0.8 nm). The outputs of all eight laser channels are passively multiplexed onto a fiber using an optical coupler and then modulated simultaneously with a 40-Gb/s NRZ-OOK signal that carries a $2^{15} - 1$ PRBS. A single commercial 40-Gb/s $LiNbO_3$ amplitude modulator is utilized, which is driven by a 40-Gb/s RF signal that is generated using a high-speed 40-Gb/s pattern generator. The multiwavelength channels are then passed through a 1.5-km span of SMF-28 to decorrelate the data and subsequently to an external SOA for packet gating.

The control header signals are created independently using three CW-DFB laser

sources at the suitable wavelengths for the frame (1555.75 nm (C27)), and two switching fabric address bits for the two-stage topology (1531.12 nm (C58), and 1543.73 nm (C42)). The control and multiwavelength payload channels are then gated into 32μsecond long packets using a DTG. The DTG act as a programmable electronic pattern generator and is synchronized with the 40-Gb/s clock. The address bits are encoded appropriately high or low for correct switching through the fabric. The channels are then multiplexed together using a passive combiner, yielding wavelength-striped optical packets including three control bits and eight 40-Gb/s data streams.

The wavelength-striped optical messages are switched within the fabric and correct path routing is verified. At the output of the switching fabric, one 40-Gb/s payload stream is filtered using a ROADM, an EDFA, a tunable filter, a VOA, then a high-speed 40-Gb/s receiver with *p-i-n* photodiode and TIA. A limiting amplifier is also utilized with two differential output ports. One of the ports is connected to an electrical demultiplexer, which time-demultiplexes the signal such that the BER can be evaluated using a 10-Gb/s BERT. The DTG is used to gate the BERT to measure the errors over the duration of the packet. No clock recovery is performed in this system, and a common clock synchronizes the DTG, pattern generator, BERT, and electrical demultiplexer.

The other differential output of the LA is connected to TiSER, which is realized using a similar platform as previous implementations. The MLL is identical, and the suitable 40-Gb/s components (*i.e.* the MZ modulator and PD) are used to support 40-Gb/s bit rates. An increased stretch factor is realized using the appropriate DCF spans; the two fiber spans are -20-ps/nm and -1310-ps/nm DCFs, respectively, realizing

a stretch factor of ∼70. Less dispersive fiber is used here (as compared to previous implementations) to avoid the dispersion penalty, which arises from low-pass filtering due to the 40-Gb/s signal sidebands' interference from dispersion. The back-end PD used here had the same specifications as previously.

TiSER is used to monitor a single 40-Gb/s channel at the output of the switching fabric. Figure 6.7 shows a photograph of the TiSER chassis as it was inserted in the switching fabric test-bed. The data is sampled using a commercial A/D digitizer with 2-GHz bandwidth, capturing up to 20 GSamples/s.

6.6.2 Results

TiSER allows the QoT of an egressing optical packet to be evaluated offline using advanced signal processing techniques. At the output of the switching fabric within the CLB, the QoT of a high-bandwidth optical packet is determined by assessing one of the 40-Gb/s optical payload channels. A sufficient number of samples is obtained to generate a 40-Gb/s eye diagram from a single optical packet. Using the sampled eye, the BER is then estimated by a calibrated signal processing algorithm that rapidly determines the resulting Q factor. In the future, a circuit board with on-board FPGA and low-speed A/D can also be used to enable the real-time, online BER extrapolation. This will allow the real-time estimation of the packets' QoT that is more rapid than a traditional BERT. Another key feature is the packet-scale BER extrapolation, which will then be leveraged in the cross-layer infrastructure to denote the optical signal quality with a packet rate.

Here, the TiSER scope is used to monitor the egression of optical packets from the

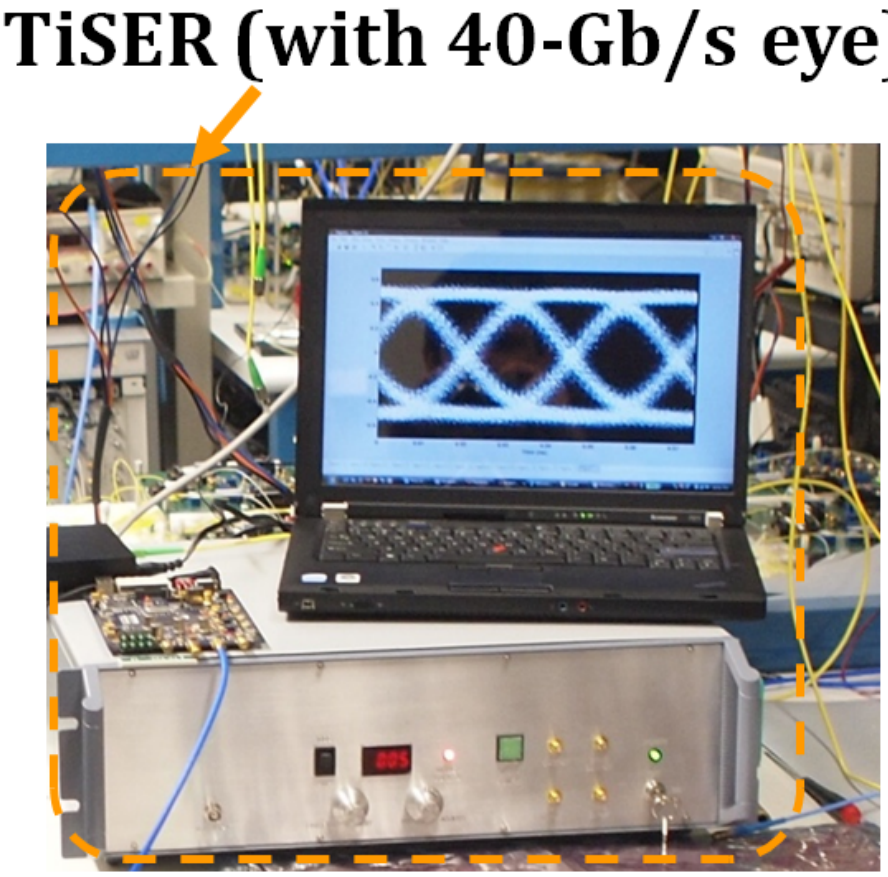

Figure 6.7: TiSER Photograph - Photograph showing the implemented TiSER scope chassis that has been implemented in the CLB.

CLB's switching fabric and allows the observation of the fabric's fast reconfiguration. A FPGA control plane can inform the fabric of a router failure or degraded link; the cross-layer OCP can then signal the switching fabric to switch routes to protect the optical packet transmission and avoid the point of failure. In this way, the packet stream can be rerouted around the failed or degraded link. The monitoring and fabric recovery capability utilizes the high-data-rate (40-Gb/s) payload channels, and the signal from the higher-layer router to the control plane is by means of a manual adjustment of a flip switch on the FPGA circuit board.

TiSER is connected one of the output ports of the switching fabric (out0). Figure 6.8 depicts the reconfiguration experiment state of an online router. Using the low-speed digitizer realized with TiSER, the optical packet stream is seen to be transmitted to the desired router link (out0). Correspondingly, Figure 6.9 depicts the reconfiguration experiment state of an offline router. The TiSER digitizer's output then thus displays no packets, since they are rerouted to alternate port (out1) within the switching fabric to avoid the packet loss of transmitting to a failed/degraded link.

Figure 6.10 shows the 40-Gb/s eye diagrams of a single optical packet (at $\lambda = 1538.98$ nm) as captured by TiSER during the fabric reconfiguration experiment. Figure 6.10a depicts the 40-Gb/s TiSER-measured eye diagram at the fabric port corresponding to the router (out0) in the case that the router is online. Figure 6.10b depicts the 40-Gb/s TiSER-captured eye diagram at the rerouted fabric port. When the router is offline or the following link is shown to be degraded, the cross-layer platform signals the optical packets to be redirected to an available output in the switching fabric (here, out1). BER estimation algorithms also show that the rerouted packets

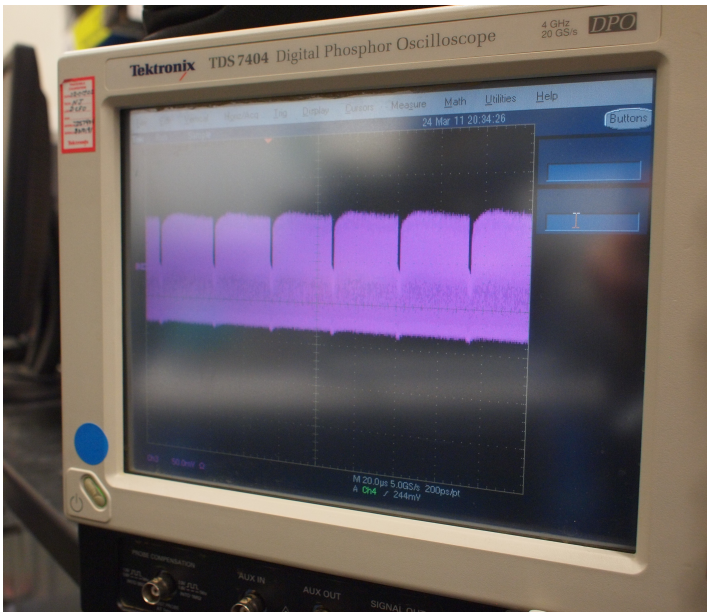

Figure 6.8: Online Router: Packet Flow - Photograph showing the optical packet flow egressing at its desired port in the case of an online router.

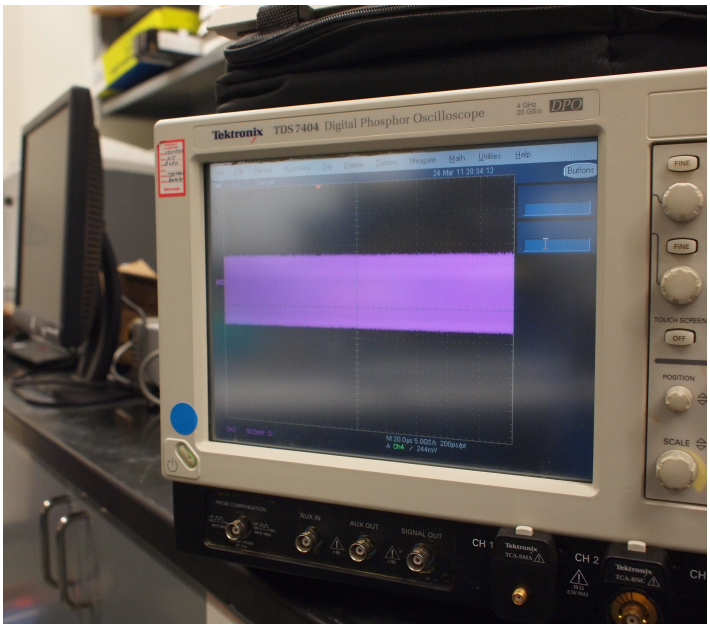

Figure 6.9: Offline Router: No Packets - Photograph showing the lack of packets egressing at their desired port in the case of an offline router.

exhibit better BER performance in the offline router case, as compared to the online router scenario.

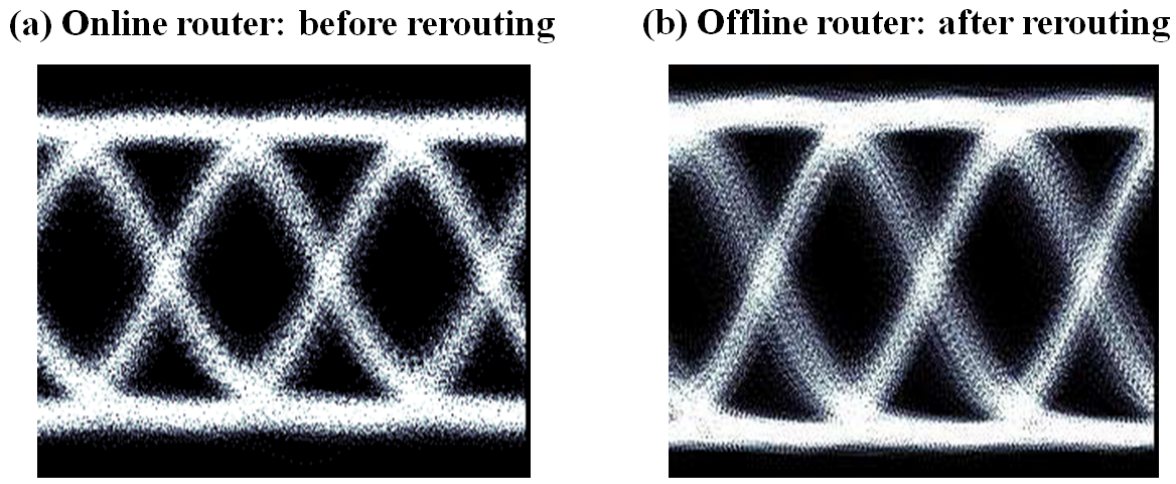

Figure 6.10: TiSER-Captured 40-Gb/s Optical Eye Diagrams - 40-Gb/s eye diagrams for the cases of an (a) online router and (b) offline router (λ = 1538.98 nm).

Using the packet analysis system outlined above, BER measurements with a commercial BERT show that all packets are switched through the fabric with error-free performance on all eight payload wavelength channels (as defined by achieving BERs less than 10^{-12}).

In order to truly demonstrate TiSER's capacity to enable the fast BER estimation at a packet rate, 40-Gb/s sensitivity curves are then obtained using TiSER alone instead of the BERT. This allows the BER measurements to be performed much faster than in previous experiments, since low BER values can be obtained without waiting long bit-checking and error-counting times. TiSER samples the data at varying optical power levels, and offline signal processing techniques are then used to estimate the Q factor. The resulting BER measurements are then plotted with respect to the

received power, similar to previous BERT-enabled experiments. Figure 6.11 shows the sensitivity curves resulting from the TiSER measurements; a ∼1.3-dB power penalty is obtained for the complete system.

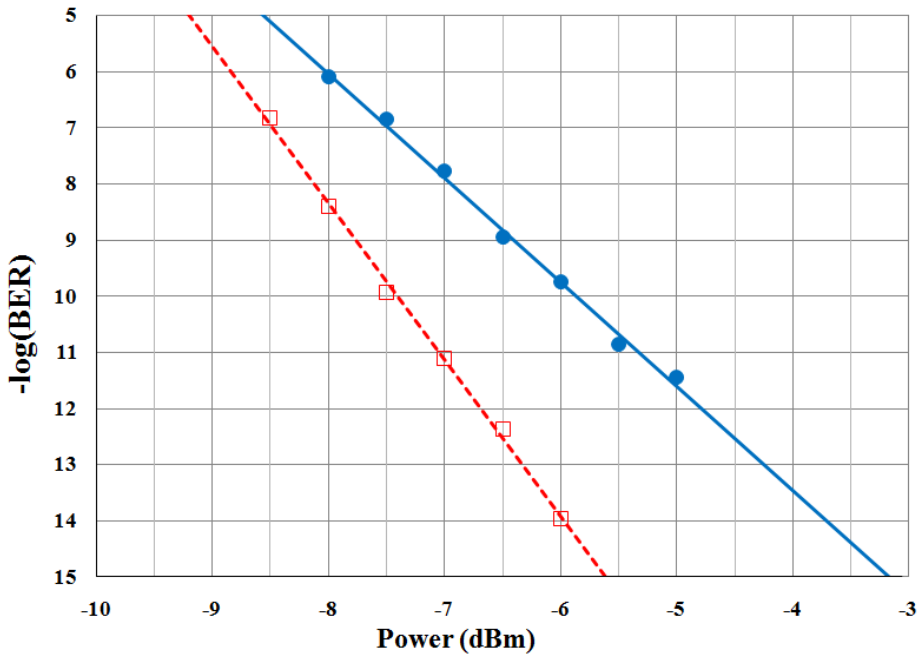

Figure 6.11: TiSER-Captured Sensitivity Curves - 40-Gb/s BER sensitivity curves for one payload channel in the online router scenario (the red, dashed line refers to the back-to-back measurements at the fabric input; the blue, solid line refers to measurements at the router output port, out0 (λ = 1538.98 nm).

This first part of the experimental demonstration truly highlights the capability of the switching fabric to quickly recover and reconfigure in the face of failures while supporting multi-terabit traffic. TiSER is used as the embedded performance monitor, showing its unique rapid BER measurement capabilities that can be used at the packet-rate. The demonstration of TiSER to monitor the 40-Gb/s channels allows the fast

measurement of the optical QoT at lower cost as compared to using commercially-available components. Here, a high-speed 40-Gb/s BERT is not required to evaluate the BER performance. Additionally, TiSER measurements can be realized at a packet-by-packet basis, which is important for future cross-layer platform designs.

6.7 Fabric Reconfiguration of HD Video Transmission

The previous section discussed the potential high capacity of the switching fabric, allowing a performance monitor to actuate a fast, packet-rate reconfiguration of 8×40-Gb/s traffic. The following section showcases the transmission of "real" data traffic via the support of a HD video streaming application, which is supported simultaneously to the high-speed PRBS data operation. As emphasized in Chapter 1, the increasing bandwidth demands are greatly driven by the network's need to enable rich multimedia applications such as real-time video.

Here, a custom-designed 10GE-based O-NIC is utilized to support the transmission of Ethernet-based video traffic through the CLB's switching fabric without distortion or frame loss. In dynamic response to router failure or optical link impairments, the cross-layer FPGA control plane allows for the switching fabric to reconfigure with a nanosecond timescale. This allows the video data to be recovered and to be transmitted seamlessly upon restoration of the optical network link. The ability to support cross-layer interactions between the application and physical layers is also demonstrated using a VBR operation of the data switched by the fabric.

6.7.1 Experimental Setup

The green region in Figure 6.6 depicts the setup for generating the 4×3.125-Gb/s wavelength-striped video streams as supported by this part of the demonstration. The O-NIC features commercial 10GE network interface cards (NICs) in the two computer end nodes (host1 and host2), connected by Quad Small Form-factor Pluggable (QSFP) cables. The NICs are connected to high-speed FPGAs and the system supports four separate lanes of 8b/10b-encoded 3.125-GBaud signals. A detailed implementation overview can be found in [175]. The O-NIC produces 4×3.125-Gb/s Ethernet-based video streams end-to-end.

Four CW-DFB lasers at the following payload wavelength channels: 1548.51 nm (C36), 1547.72 nm (C37), 1546.92 nm (C38), and 1546.12 nm (C39), are used to create the optical link. The Ethernet data is generated by the source host, which drives four $LiNbO_3$ modulators. The multiwavelength data is then multiplexed with the appropriate control headers and injected in the CLB's fabric. Circuit-switched paths are established for the video streams. At the output of the fabric, the data is appropriately filtered and received using four *p-i-n* receivers with TIA and LA pairs, and transmitted to the destination host. The upstream traffic is looped back electronically. The photographs in Figure 6.12 show some of the optical components implemented by the O-NIC.

6.7.2 Results

The O-NIC is used to demonstrate HD video streaming over the two-stage switching fabric test-bed. The video is observed to be transmitted without distortion or the

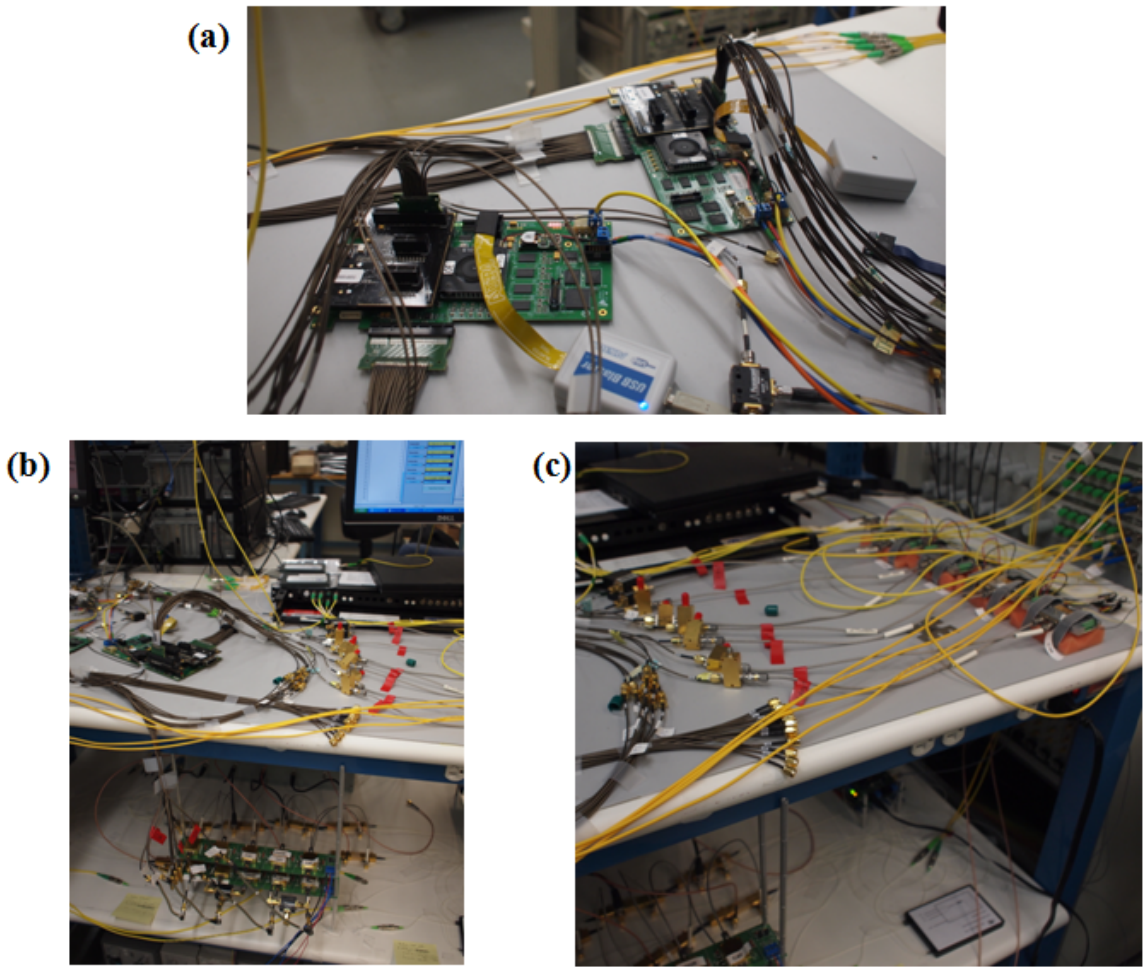

Figure 6.12: O-NIC Photographs - Photographs of the O-NIC, showing the (a) FPGA circuit boards, (b) four $LiNbO_3$ modulators, and (c) four *p-i-n* receivers.

loss of frames. Figure 6.13 presents a photograph of the two host computer screens, showing one host playing a recorded video through the 10GE-based optical network link on the other host.

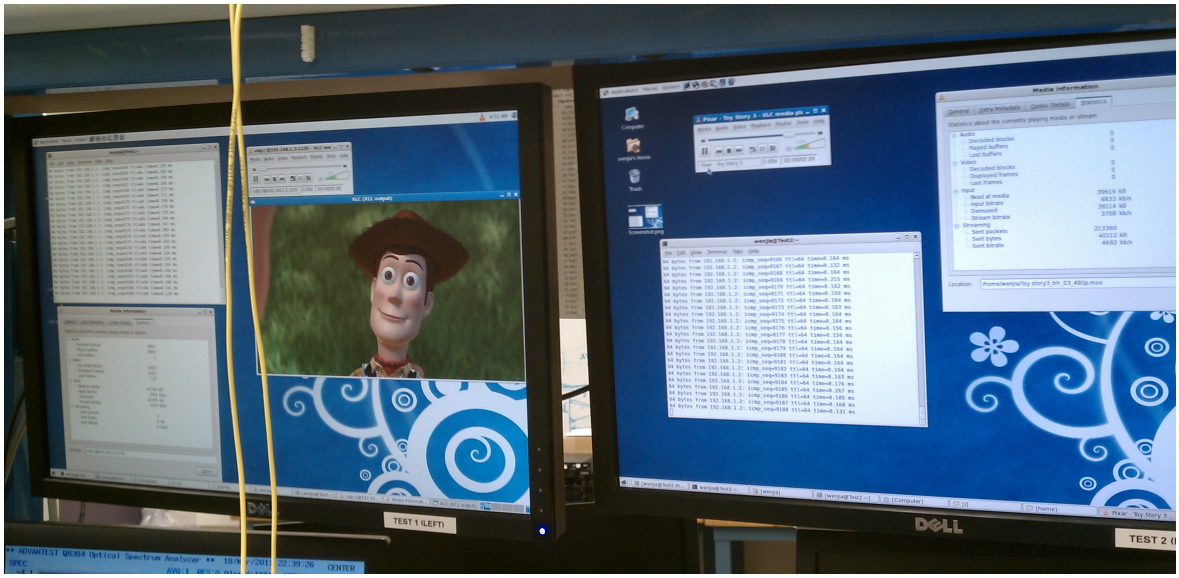

Figure 6.13: Video Streaming - Photograph of the video streaming demonstration, showing the two hosts' monitors as one computer is transmitting video to the other.

The cross-layer reconfiguration is again shown for the video streaming in which the control plane can signal the switching fabric to reroute the optical packets in the detection of an optical link degradation. During the lightpath rerouting, the video is paused for a short time while the Ethernet link is restored, due the lack of burst-mode operability of the O-NIC transceivers.

Further, in order to demonstrate the cross-layer adaptability of the application layer with the optical physical layer, a VBR transmission is set up over the CLB's switching fabric. The two host computers that are connected through the optical packet switch leverage the 10GE interface described above, effectively creating a two-host private IP

network. The destination (host2) displays the images originating from a HD webcam physically connected to the source (host1); the video is encoded using software based on FFmpeg [176] and streamed in the form of User Datagram Protocol (UDP) packets. Figure 6.14 shows the real-time streaming-over-optics of webcam images of several authors involved in this work.

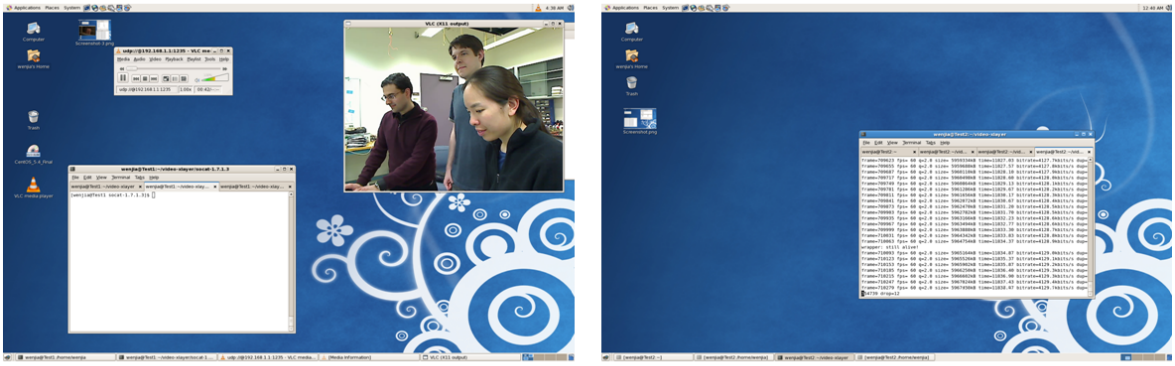

Figure 6.14: Webcam Streaming - Screenshots of the two hosts, showing the streaming of HD webcam images of the authors of this work from the source host to the destination host.

The video encoded is customized such that the codec parameters could be modified on-the-fly. The system switches between high bit rates (supporting high-quality video) and degraded bit rates (supporting low-quality video) upon receiving signaling commands embedded in specific UDP packets. The signals are sent from host2 (destination) to host1 (source); in the future, this could be carried using out-of-band signaling to another network interface on the source host.

Figure 6.15 shows the screenshots of the high-quality and low-quality video images as enabled by the VBR demonstration.

Figure 6.15: Variable-Bit-Rate Demonstration - Screenshots of the VBR transmission, showing the support for low-quality video streams (left) and high-quality video data (right).

In this particular demonstration, the cross-layer signaling was manual, whereby the control UDP packets are sent by user command. In a practical networking scenario, various performance monitoring subsystems can detect the QoT degradations and/or increases in BER on a link, and subsequently signal the control plane. The OCP can then instruct the transponders at the sending and/or receiving ends to reduce the link's bit rate for improved impairment resiliency, and inform the higher-layer application layer of these changes to allow for the network to cope with reduced resources.

A block schematic of the second part of the demonstration is provided in Figure 6.16, highlighting how the experimental results are envisioned to fit into the overall network architecture. The various PM subsystems are shown to feedback to the FPGA-based control plane which can then dynamically control the CLB. The CLB nodes exhibit various functionalities; the high- and low-quality videos are the possible result of

optical-layer impairments.

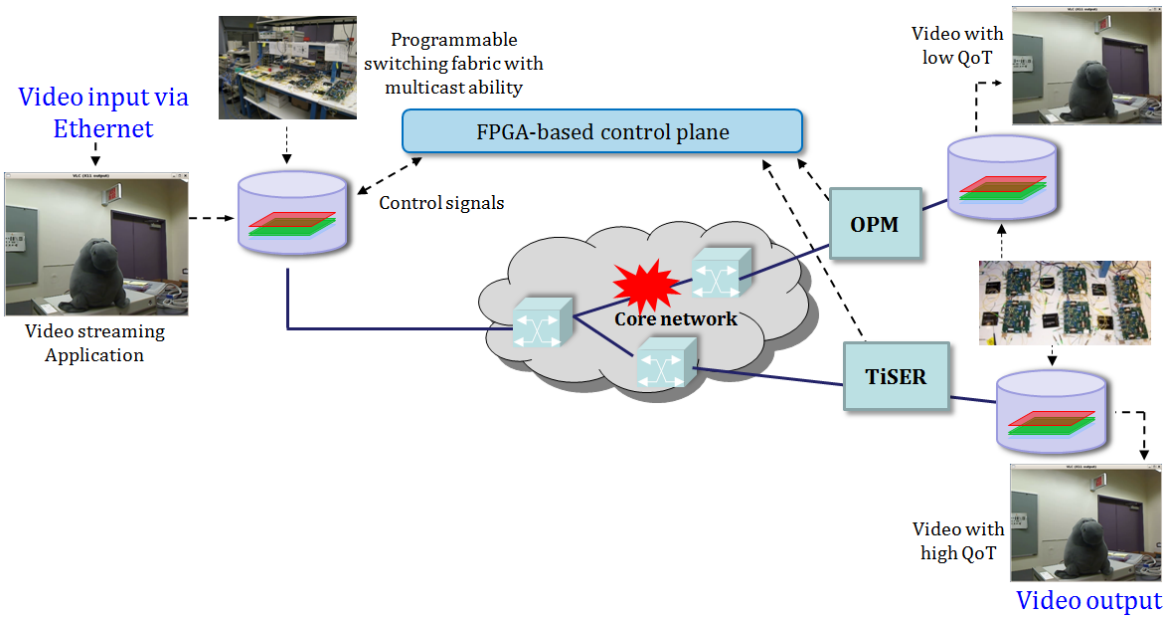

Figure 6.16: Demonstration with Results - High-level block diagram of the video transmission demonstration.

Future experimental realizations of the cross-layer box can further explore the issue of optical data aggregation, as well as leverage past work on enabling advanced switching fabric functionalities (*e.g.* use other optical performance monitoring modules, support QoS-based optical switching, optical packet multicasting, *etc.*)

6.8 Collaborations with GENI

With the help of the author, there has been an initial deployment of a stripped-down version of the cross-layer box within an existing optical network test-bed in GENI. In [52], a NetFPGA-based [177] cube has been realized in the Breakable Experimental Network (BEN) [178] test-bed. BEN is a four-node ring network geographically located

in North Carolina. The SILO architecture, which has been developed by collaborators in [51], is also implemented here as a forward-looking software framework that supports cross-layer tuning and control via the use of cross-service knobs.

A high-level overview of the experiment is as follows: optical power fluctuations are introduced using a VOA that is placed inline between two BEN network nodes. The power changes are detected using an optical cross-connect with power monitoring capabilities, which can then feedback to the higher-layer services and inform a NetFPGA-based control plane. The NetFPGA interfaces with an inline optical amplifier (*e.g.* a SOA) which can be gated on to dynamically compensate for the power fluctuations.

A QoS-aware streaming video application is used to exemplify the cross-layer scheme. The power changes for the high-QoS video stream are compensated for using the SOA, while the power fluctuations for the low-QoS stream remains unchanged. Using software modifications, the power variations can be mapped onto the video signals such that no degradations can been visually observed for the high-QoS stream.

More detailed information regarding this demonstration can be found in [52]. The work shows the backing from the GENI and networking communities regarding the need for a cross-layer architectural design for future optical networks.

6.9 Closing Remarks

As a final capstone of this dissertation, this chapter presents the cumulation of the author's work, providing the design and demonstration of the first prototype of a cross-layer aggregation box. The plan is for the CLB to be inserted in the mesh

access/aggregation network to provide physical-layer awareness, fast optical switching, and real-time reactive routing. The intelligent resource allocation that it will support can then deliver high-bandwidth capacities in real-time, with low cost and low energy.

The box's functionalities and capabilities will undoubtedly evolve in future versions of prototypes, particularly with respect to the real-time reaction to the optical link's status/performance. The current implementation of the box leverages emerging device technologies from other CIAN-related research groups, as well as commercial off-the-shelf components; as the box evolves, a greater number of innovative optical technologies will be utilized in the box. Additionally, the box will continue to incorporate cross-layer capabilities with higher-layer protocols and algorithms. In the future, the CLB will perform a holistic switching co-optimization with inputs consisting of the incoming data's QoS, the physical-layer QoT, and energy, to result in an intelligent optical packet aggregation node.

This work constitutes a fundamental stepping stone to realizing future systems-level endeavors of creating integrated network elements that support optical switching and packet-scale reconfiguration. The cross-layer enabled node ensures that high optical QoT constraints are satisfied while taking into account multiple QoS and energy requirements from the higher layers. This specific effort is extremely important to fulfilling CIAN's mission of inserting multiple CLBs in future mesh-based access/aggregation networks.

Chapter 7

Summary and Conclusions

This final chapter concludes the dissertation, presenting a glimpse at some of the current and future research projects at the cross-layer test-bed environment that has been pioneered, established, and developed by the author of this thesis.

Lastly, some concluding remarks on the contributions of this dissertation are provided.

7.1 Global Picture

This work has been motivated by the requirement to support extremely agile, highly-functional, and reliable optical connectivity without excessive overprovisioning in future networks. In order to enable these intelligent optical networking functionalities at low cost and with extreme energy efficiency, the author envisions an advanced cross-layer infrastructure that allows optical switching to be executed at a packet-timescale incorporating inputs and performance metrics from all layers of the OSI stack. The goal is to create seamless, more transparent paths across an intelligent network that

can dynamically deliver the most demanding high-bandwidth applications.

This cross-layer world view is shared by numerous top researchers and networking firms in industry, as well as by many collaborators in the optical research community (including leaders such as Kilper, Willner, Winzer *etc.*) In [179], Ciena claims that a multi-layer approach with enhanced network intelligence – similar to this cross-layering concept – can provide improvements in bandwidth allocation, network management, as well as minimize the overall switching and transmission costs. An advanced level of intelligent automation allows the multi-layer control plane to discover and manage all photonic and higher-layer resources, to handle lightpath selection and packet routing, and to realize traffic grooming, while simultaneously meeting service requirements (*i.e.* QoS) and co-optimizing energy and QoT.

As emphasized in this thesis, the ultimate goal is a high-performance optical network node (the CIAN Box). The box is an intelligent network element with distributed control plane management, cross-layer capabilities, multi-layer traffic engineering, fast optical switching, and packet-level monitoring embedded in the physical layer. It provides enhanced awareness of the physical optical channels, as well as the ability to react to the physical-layer awareness on a packet-timescale and the support of cross-layer reaction. With the endeavor of achieving advanced multi-layer routing and control algorithms, the network node requires an intelligent co-optimization across all the layers.

7.2 Future Work

There are many ongoing research activities that have been initiated by the author and that will continue beyond the time of writing of this thesis. The advanced networking functionalities of the optical switching fabric will continue to be developed by others (with the help of the author), including a validation of the switching fabric's modulation-format transparency.

Other planned work at the cross-layer test-bed includes the construction of a small-scale network of CIAN Boxes. The network will have a basic mesh topology, in accordance with the vision of inserting the CIAN Box in future access/aggregation networks complementing industry's evolution towards a mesh-centered design. The test-bed will be comprised of various prototypes of CLBs, featuring the dynamic packet-rate reconfiguration times and distributed control capabilities. This allows the core-edge interface to be physical-layer aware with real-time reactive routing.

Future CLB prototypes will incorporate the reprogrammable optical switching fabric, extensive dynamic packet-level optical performance monitoring capabilities, other complementary mature CIAN-developed optical devices, as well as support an optical/wireless interface. The test-bed will be utilized to carry out various cross-layer experiments and to validate advanced functionality hypotheses, particularly from the algorithmic and routing protocol perspective. The CLBs' support of typical and emerging wireline and wireless protocols will be a result of collaborations with Zussman *et al.*

To this end, the CLB network will allow for the demonstration of the following functionalities (among others):

- HD video transmission;
- physical-layer reconfiguration of a network of CLB nodes following IP-layer failure or router sleep cycles; and
- real-time optical performance monitoring and cross-layer signaling with the application layer.

The CIAN CLBs constitute prototypes of the forward-looking integrated optical aggregation nodes. The box will be capable of real-time monitoring and on-the-fly reconfiguration, fully controlled by a general-purpose computer that will have access to live network traffic. Multiple CLBs will be constructed to showcase the support of high-bandwidth applications (*i.e.* QoS-aware HD video streaming) on the small-scale experimental network. The resulting platform will be capable of implementing various dynamic routing algorithms and multiple cross-layer applications, in conjunction with optical monitoring solutions, while communicating with a computer to expose the required application programming interface (API) to enable impairment-aware applications.

Additionally, one goal of the future version of the CLB is to possess the ability to support high-bandwidth end-to-end connectivity between edge users for a fixed duration (*i.e.* in the situation of a teleconference); the high-capacity link can then be torn down when it is no longer required. It is widely recognized that future telecom networks should be able to schedule these broadband applications "on demand" as required [180]. The intelligent resource allocation arising from the deployment of the CIAN Box can then facilitate delivering high-bandwidth applications at low cost and

in real-time.

Therefore, the following capabilities will be supported by future CIAN CLBs:

- a cross-layer optimized platform,
- fast programmable optical switching on the packet rate,
- dynamic performance measurements,
- energy-aware optimization protocols and algorithms,
- delivery of multiple QoS classes, and
- heterogeneous traffic.

Currently, CIAN's vision is that these functionalities can be achieved within the goals of reducing the energy consumption, fast optical switching, and increased efficiency in bandwidth utilization. The impact of the ubiquitous deployment of CIAN boxes throughout the network will be the following key metrics and benchmarks:

- reduce energy consumption from the current state-of-the-art technologies by ∼10s×, by means of optical switching, sleep modes, cross-layer coordination, *etc.* , working with Bell Labs, Alcatel-Lucent under GreenTouch;
- enable faster, finer (per-packet) granular data to achieve per-packet optical reconfigurability, via the integration of forward-looking CIAN-driven devices; today's state of the art is on the order of hundreds of milliseconds, and here the goal is ∼tens of nanoseconds;

- increase bandwidth utilization via dynamic resource allocation via OPM modules and QoT support, using QoS from the higher layers; bandwidth utilization in today's networks is less than 20%, and the goal is to improve this metric by 4×.

Future capabilities of the CIAN Box include aggressive goals of realizing throughputs greater than 100 Tb/s, with a total power consumption on the order of 1 kW. Per-packet reconfiguration times on the order of 10s of nanoseconds are also within the projected aims of following successive CLB prototypes.

Future OPM capabilities in the cross-layer test-bed will also be expanded to encompass a greater set of optical modulation formats, at higher data rates, with the real-time introspection infrastructure capable of extracting a greater number of performance monitoring metrics (*e.g.* PMD, CD, *etc.* , in continuing collaborations with Willner *et al.*). Performance monitoring collaborative work to enable the real-time extrapolation of the data's BER will also continue with Jalali *et al.* The goal of these introspective technologies is to be able to extract these measurements in real-time and actuate advanced switching protocols on the optical layer.

Furthermore, the current discrete-component implementation of the switching fabric will be replaced by integrating numerous research devices that are currently being developed by the aid of the author and in collaboration with other major CIAN institutions. For example, these will be the result of pending collaborative efforts with Fainman *et al.* at the University of California, San Diego (UCSD), Peyghambarian and Norwood *et al.* at the University of Arizona (UA), *etc.* Instead of using commercially-available parts, these emerging optical components and technologies will be inserted in the test-bed to allow for a far more integrated, cost-effective, and potentially energy-

efficient realization of the optical switching fabric for future CLB prototypes.

Additionally, in order to evaluate the performance of the algorithms in a realistic environment, a wireless mesh network will be integrated with the switching fabric test-bed (work to be performed with Zussman *et al.*). 802.11 access points will be connected to the optical switching fabric, with the goal of creating wireless-optical bridges. Figure 7.1 depicts the network layers that will be required to set up a wireless bridge/interface point in the test-bed. In the long term, the envisioned wireless-optical environment will be integrated with emerging wireless standards (*e.g.* WiMAX) through the integration of a WiMAX node that is currently being deployed at Columbia as part of GENI [181]. WiMAX is a wireless 4G technology that is currently being deployed by industry in major US cities. 4G technologies (*i.e.* WiMAX and LTE) are expected to aggravate the backhaul load on the aggregation network, which will further challenge future networks. Through this integration at the cross-layer test-bed, various cross-layer backhauling algorithms can then be tested and validated on an industry-accepted platform, investigating the effect of these algorithms under a high-volume wireless traffic.

7.3 Final Thoughts

To close the dissertation: the design of the next-generation Internet infrastructure is driven directly by the need to address the massive growth in bandwidth demands and network traffic, in addition to the challenges arising from the unsustainable acceleration in energy consumption growth. The bottleneck in delivering efficient, low-cost, high bandwidths to a multitude of users and heterogeneous applications is the principal

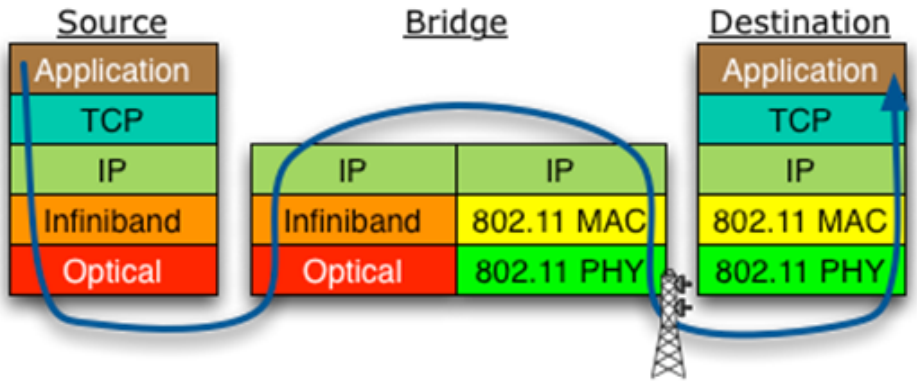

Figure 7.1: Wireless-Optical Bridge - Network stack for enabling a wireless-optical interface.

driver for this work. Therefore, the research community must adopt radically novel, energy-efficient, optical networking technologies and architectures to truly sustain the explosive growth in user demand.

This dissertation aims to tackle the critical challenge of designing and implementing a unifying architectural platform that allows cross-layer optimization directly with the optical layer. The cross-layer platform endeavors to provide a new framework for future networks to incorporate packet-level measurement techniques, schemes for monitoring the health of optical channels, and performance prediction in next-generation multi-terabit networks. Allowing a more intelligent programmable optical layer can then achieve greater flexibility with respect to bandwidth allocation and potentially a significant reduction in the network's energy consumption.

The important take-home point is that the cross-layer design should utilize, drive, and innovate optical technologies and systems in a novel way, bridging the gap between the development and fabrication of basic optical devices and the algorithms employed

by the higher network layers. The primary driver is that the performance of optical components and routing applications should be fine-tuned simultaneously and unison, to dynamically and holistically optimize the optical physical layer and the networking layers in concert. This advancement in cross-layer network design transforms the high-bandwidth optical "pipe" to an intelligent traffic delivery system that is also extremely flexible and intelligently aware of its higher layers.

Deploying these agile, intelligent optically switched nodes may face significant adoption hurdles moving forward in the next phases of redesigning the Internet; however, it has great potential to realize highly dynamic, energy-efficient, and high-capacity communication links in the next-generation Internet and other optical networking infrastructures.

Glossary

10GE 10-Gigabit Ethernet

3D Three dimensional

A/D Analog-to-digital

API Application programming interface

ASE Amplified spontaneous emission

AWG Arrayed waveguide grating

BEN Breakable Experiment Network

BER Bit-error rate

BERT Bit-error-rate tester

CapEx Capital expenditures

CD Chromatic dispersion

CIAN Center for Integrated Access Networks

CLB Cross-layer box

CPLD Complex programmable logic device

CRC Cyclic redundancy check

CSA Communications signal analyzer

CW Continuous wave

DCA Digital communications analyzer

DCF Dispersion-compensating fiber

DFB Distributed feedback

DICONET Dynamic Impairment Constraint Networking for Transparent Mesh Networks

DLI Delay-line interferometer

DPSK Differential Phase-Shift Keying

DQPSK Differential Quaternary Phase-Shift Keying

DTG Data timing generator

EDFA Erbium-doped fiber amplifier

FDL Fiber delay line

FEC Forward error correction

FIFO First-in first-out

FPGA Field-programmable gate array

FRR Fast reroute

FSR Free spectral range

GENI Global Environment for Network Innovations

GMPLS Generalized Multi Protocol Label Switching

GPIO General purpose input/output

HD High definition

HPC High-performance computing

ICT Information and Communication Technologies

IP Internet Protocol

ITU International Telecommunication Union

LA Limiting amplifier

$LiNbO_3$ Lithium niobate

MDR Minimum distance routing

MHR Minimum hop routing

MINTS Minnesota Internet Traffic Study

MPMA Multistage Packet Multicasting Architecture

MTU Maximum transmission unit

MZI Mach-Zehnder interferometer

NIC Network Interface Card

NRZ Nonreturn-to-zero

O-NIC Optical Network Interface Card

O/E/O Optical/electronic/optical

OBS Optical burst switching

OCP Optical control plane

OCS Optical circuit switching

OFS Optical flow switching

OIN Optical interconnection network

OOK ON-OFF-keyed

OpEx Operational expenditures

OPM Optical performance monitoring

OPS Optical packet switching

OQoS Optical quality of service

OSA Optical spectrum analyzer

OSI Open Systems Interconnection

OSNR Optical signal-to-noise ratio

PaM Packet multicasting

PaR Packet routing

PC Polarization controller

PD Photodetector

PG Pattern generator

PIC Photonic integrated circuit

PIN *p-i-n* photodetector

PM Performance monitoring

PMD Polarization-mode dispersion

PPG Pulse pattern generator

PPT Proactive packet protection

PRBS Pseudo-random bit sequence

PSaD Packet Splitter and Delivery

PSE Photonic switching element

Q Quality (factor)

QoS Quality of service

QoT Quality of transmission

QSFP Quad Small Form-factor Pluggable

RAM Random access memory

RBS Real-time burst sampling

RF Radio frequency

ROADM Reconfigurable optical add-drop multiplexer

RWA Routing and wavelength assignment

RZ Return-to-zero

SILO Services Integration, controL, and Optimization

SMF Single-mode fiber

SOA Semiconductor optical amplifier

SONET/SDH Synchronous Optical Networking/Synchronous Digital Hierarchy

TDM Time-division multiplexing

TIA Transimpedance amplifier

TiSER Time-stretch enhanced recording

UDP User Datagram Protocol

VBR Variable-bit-rate

VOA Variable optical attenuator

WDM Wavelength-division multiplexing

References

[1] L. Kleinrock, "On communications and networks," *IEEE Transactions on Computers*, vol. C-25, no. 12, pp. 1326–1335, 1976. 1

[2] A. Odlyzko, "Minnesota Internet Traffic Studies," [Online]: http://www.dtc.umn.edu/mints/. 2

[3] R. W. Tkach, "Scaling optical communications for the next decade and beyond," *Bell Lab. Tech. J.*, vol. 14, pp. 3–9, February 2010. 2, 3, 6, 81

[4] P. J. Winzer, "Challenges and evolution of optical transport networks," in *36th European Conference and Exhibition on Optical Communication (ECOC)*, 2010, paper We.8.D.1. 2, 6, 81

[5] J. D'Ambrosia, "40 gigabit Ethernet and 100 gigabit Ethernet: The development of a flexible architecture [commentary]," *IEEE Communications Magazine*, vol. 47, no. 3, pp. S8–S14, 2009. 2

[6] "Cisco Visual Networking Index: Forecast and Methodology (2009-2014) [White Paper]," Cisco, Tech. Rep., 2010, [Online]: http://www.cisco.com/. 2, 4, 6

[7] B. Swanson and G. Gilder, "Estimating the Exaflood: The impact of video and rich media on the internet a zettabyte by 2015?" Discovery Institute, Tech. Rep., 2008, [Online]: http://www.discovery.org/a/4428. 3

[8] D. W. Schloerb, "A quantitative measure of telepresence," *Presence*, vol. 4, no. 1, pp. 64–80, 1995. 5

[9] Panasonic Life Wall, [Online]: http://panasonic.net. 5

[10] P.-A. Blanche, A. Bablumian, R. Voorakaranam, C. Christenson, W. Lin, T. Gu, D. Flores, P. Wang, W.-Y. Hsieh, M. Kathaperumal, B. Rachwal, O. Siddiqui, J. Thomas, R. A. Norwood, M. Yamamoto, and N. Peyghambarian, "Holographic three-dimensional telepresence using large-area photorefractive polymer," *Nature*, vol. 468, pp. 80–83, 2010. 5

[11] "2010 Global Broadband Phenomena: Research Report," Sandvine, Tech. Rep., 2010, [Online]: http://www.sandvine.com/. 6

[12] S. Elby, "The future internet - a service provider's long term view," in *IEEE/LEOS Summer Topical Meeting (LEOSST)*, 2009, paper TuD1.1. 6

[13] G. Raybon and P. J. Winzer, "100 Gb/s challenges and solutions," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2008, paper OTuG1. 6

[14] R. J. Shapiro, "The internet's capacity to handle fast-rising demand for bandwidth," US Internet Industry Association, Tech. Rep., 2010, [Online]: http://www.usiia.org/. 6

[15] J. D'Ambrosia, "100 gigabit ethernet and beyond [commentary]," *IEEE Communications Magazine*, vol. 48, no. 3, pp. S6 –S13, 2010. 6

[16] G. P. Agrawal, *Fiber-Optic Communication Systems*, 3rd ed. New York, NY, USA: John Wiley & Sons, Inc., 2002. 7, 62

[17] K. C. Kao and G. A. Hockham, "Dielectric-fibre surface waveguides for optical frequencies," *Institution of Electrical Engineers,*, vol. 113, no. 7, pp. 1151–1158, 1966. 7

[18] R. J. Mears, L. Reekie, I. M. Jauncey, and D. N. Payne, "Low-noise erbium-doped fibre amplifier operating at 1.54μm," *Electronics Letters*, vol. 23, no. 19, pp. 1026–1028, 10 1987. 7

[19] J. Berthold, A. A. M. Saleh, L. Blair, and J. M. Simmons, "Optical networking: Past, present, and future," *Journal of Lightwave Technology*, vol. 26, no. 9, pp. 1104–1118, May 2008. 7

[20] NSF Engineering Research Center for Integrated Access Networks (CIAN), [Online]: http://cian-erc.org/. 7, 12, 13, 169, 204

[21] C. R. Doerr, "High performance photonic integrated circuits for coherent fiber communication," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2010, paper OWU5. 7

[22] P. J. Winzer, "Beyond 100G ethernet," *IEEE Communications Magazine*, vol. 48, no. 7, pp. 26–30, 2010. 7, 90

[23] S. J. B. Yoo, "Optical packet and burst switching technologies for the future photonic internet," *Journal of Lightwave Technology*, vol. 24, no. 12, pp. 4468–4492, 2006. 8, 23, 62

[24] R. Ramaswami and K. N. Sivarajan, *Optical networks: a practical perspective*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2002. 8, 123

[25] A. Shacham and K. Bergman, "An experimental validation of a wavelength-striped, packet switched, optical interconnection network," *Journal of Lightwave Technology*, 2009. 8, 40, 75, 82, 94, 124

REFERENCES

[26] C. P. Lai, D. Brunina, and K. Bergman, "Demonstration of 8×40-Gb/s wavelength-striped packet switching in a multi-terabit capacity optical network test-bed," in *23rd Annual Meeting of the IEEE Photonics Society*, 2010, paper ThQ2. 8, 40, 42, 54, 82, 90, 124, 175, 197, 217, 218

[27] I. Baldine, "Unique optical networking facilities and cross-layer networking," in *IEEE/LEOS Summer Topical Meeting (LEOSST)*, 2009, paper TuD4.2. 9, 24, 74

[28] C. Politi, C. Matrakidis, and A. Stavdas, "Cross layer routing in transparent optical networks," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2007, paper JThA64. 9, 25

[29] K. Gammon, "Four ways to reinvent the internet," *Nature*, vol. 463, pp. 602–604, 2010. 9, 118, 120

[30] C. P. Lai and K. Bergman, "Cross-layer communications for high-bandwidth optical networks," in *12th International Conference on Transparent Optical Networks (ICTON)*, 2010, paper Tu.C3.2. 9, 120, 196

[31] "Mobile backhaul evolution and convergence [white paper]," NEC Corporation, Tech. Rep., 2010. 15

[32] F. Hao and E. Zegura, "On scalable qos routing: performance evaluation of topology aggregation," in *IEEE International Conference on Computer Communications (INFOCOM)*, vol. 1, 2000, pp. 147–156. 16

[33] K. Casier, W. Tavernier, D. Colle, M. Pickavet, D. Papadimitriou, and P. Demeester, "Forecasting cost trends for carrier ethernet," in *IEEE GLOBECOM Workshops*, dec. 2009, pp. 1–6. 16

[34] Y. K. Lize, J.-Y. Yang, L. Christen, X. Wu, S. Nuccio, T. Wu, A. E. Willner, R. Kashyap, and F. Seguin, "Simultaneous and independent monitoring of OSNR, chromatic and polarization mode dispersion for NRZ-OOK, DPSK and Duobinary," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2007, paper OThN2. 16, 141, 169, 171, 174, 180

[35] C. P. Lai, M. S. Wang, A. S. Garg, K. Bergman, J.-Y. Yang, M. R. Chitgarha, and A. E. Willner, "Demonstration of QoS-aware packet protection via cross-layer OSNR signaling," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2010, paper OTuM2. 16, 121, 141, 171

[36] A. E. Willner, J.-Y. Yang, and X. Wu, "Optical performance monitoring: Perspectives and challenges," in *Optical Performance Monitoring Techniques for Next Generation Photonic Networks*, C. Chan, Ed. Elsevier Publishers, Academic Press, San Diego, 2010, pp. 1–19. 18, 169

[37] D. C. Kilper, R. Bach, D. J. Blumenthal, D. Einstein, T. Landolsi, L. Ostar, M. Preiss, and A. E. Willner, "Optical performance monitoring," *Journal of Lightwave Technology*, vol. 22, no. 1, pp. 294–304, 2004. 18, 169

[38] A. Cianfrani, V. Eramo, A. Germoni, C. Raffaelli, and M. Savi, "Loss analysis of multiple service classes in shared-per-wavelength optical packet switches," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 1, no. 2, pp. A69 –A80, 2009. 21, 75

[39] S. Perrin, "The need for next-generation ROADM networks [white paper]," Heavy Reading, Tech. Rep., 2010. 22, 64, 88

[40] D. C. Kilper, "Energy efficient networks," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2011, paper OWI5. 22, 33, 34

[41] M. Z. Feng, K. Hinton, R. Ayre, and R. S. Tucker, "Energy efficiency in optical ip networks with multi-layer switching," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2011, paper OWI2. 22, 23

[42] O. Tamm, C. Hermsmeyer, and A. M. Rush, "Eco-sustainable system and network architectures for future transport networks," *Bell Labs Technical Journal*, vol. 14, no. 4, pp. 311–327, 2010. 22, 33

[43] W. Wei, Q. Zeng, Y. Ouyang, and D. Lomone, "Differentiated integrated QoS control in the optical internet," *IEEE Communications Magazine*, vol. 42, no. 11, pp. S27–S34, 2004. 23, 31

[44] V. Chan, "Scalable control plane architecture for optical flow switched networks," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2011, paper OWP4. 23

[45] G. Weichenberg, V. W. S. Chan, and M. Medard, "Performance analysis of optical flow switching," in *IEEE International Conference on Communications (ICC)*, 2009, pp. 1–6. 23

[46] V. W. S. Chan, G. Weichenberg, and M. Medard, "Optical flow switching," in *Broadband Communications, Networks and Systems, 2006. BROADNETS 2006. 3rd International Conference on*, 2006, pp. 1 –8. 23

[47] B. Mukherjee, "Optical networks: The road ahead [invited]," in *33th European Conference and Exhibition on Optical Communication (ECOC)*, 2007, paper 8.5.1. 23

[48] A. Stavdas, C. T. Politi, T. Orphanoudakis, and A. Drakos, "Optical packet routers: how they can efficiently and cost-effectively scale to petabits per second [invited]," *Journal of Optical Netwowrking*, vol. 7, no. 10, pp. 876–894, Oct 2008. 24, 39

[49] "Evolution to optical packet transport [White Paper]," Ericsson, Tech. Rep., 2009, [Online]: www.ericsson.com/. 24

[50] Global Environment for Network Innovations (GENI), [Online]: http://www.geni.net/. 24

[51] A. Wang, "The SILO architecture: Exploring future internet design," Ph.D. dissertation, North Carolina State University, North Carolina, 2010. 24, 235

[52] M. S. Wang, A. Wang, B. G. Bathula, C. P. Lai, I. Baldine, C. Chen, D. Majumder, D. Gurkan, G. N. Rouskas, R. Dutta, and K. Bergman, "Demonstration of qos-aware video streaming over a metro-scale optical network using a cross-layer architectural design," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2011, paper NThC4. 24, 234, 235

[53] B. Ramamurthy, D. Datta, H. Feng, J. P. Heritage, and B. Mukherjee, "Impact of transmission impairments on the teletraffic performance of wavelength-routed optical networks," *Journal of Lightwave Technology*, vol. 17, no. 10, pp. 1713–1723, 1999. 25, 169

[54] Y. Huang, J. P. Heritage, and B. Mukherjee, "Connection provisioning with transmission impairment consideration in optical WDM networks with high-speed channels," *Journal of Lightwave Technology*, vol. 23, no. 3, pp. 982–993, 2005. 25

[55] C. V. Saradhi and S. Subramaniam, "Physical layer impairment aware routing (PLIAR) in WDM optical networks: Issues and Challenges," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 4. 25

[56] S. Rai, C. F. Su, and B. Mukherjee, "On provisioning in all-optical networks: An impairment-aware approach," *IEEE/ACM Transactions on Networking*, vol. 17, no. 6, pp. 1989–2001, 2009. 25

[57] B. G. Bathula, R. R. C. Bikram, V. M. Vokkarane, and S. Talabattula, "Quality-of-Transmission-Aware manycasting over optical burst-switched networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 2, no. 10, pp. 820–830, 2010. 25, 159

[58] P. Pavon-Marino, S. Azodolmolky, R. Aparicio-Pardo, B. Garcia-Manrubia, Y. Pointurier, M. Angelou, J. Sole-Pareta, J. Garcia-Haro, and I. Tomkos, "Offline impairment aware rwa algorithms for cross-layer planning of optical networks," *Journal of Lightwave Technology*, vol. 27, no. 12, pp. 1763–1775, 2009. 25

[59] S. Azodolmolky, Y. Pointurier, M. Klinkowski, E. Marin, D. Careglio, J. Sole-Pareta, M. Angelou, and I. Tomkos, "On the offline physical layer impairment aware rwa algorithms in transparent optical networks: State-of-the-art and beyond," in *International Conference on Optical Network Design and Modeling (ONDM)*, 2009, pp. 1–6. 25, 169

[60] F. Agraz, S. Azodolmolky, M. Angelou, J. Perello, L. Velasco, S. Spadaro, A. Francescon, C. Saradhi, Y. Pointurier, P. Kokkinos, E. Varvarigos, M. Gunkel, and I. Tomkos, "Experimental demonstration of centralized and distributed impairment-aware control plane schemes for dynamic transparent optical networks," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2010, paper PDPD5. 25

[61] Y. Qin, K. C. S. Cheng, J. Triay, E. Escalona, G. Zervas, G. Zarris, N. Amaya-Gonzalez, C. Cervello-Pastor, R. Nejabati, and D. Simeonidou, "Demonstration of c/s based hardware accelerated qot estimation tool in dynamic impairment-aware optical network," in *36th European Conference and Exhibition on Optical Communication (ECOC)*, 2010, paper P5.17. 25

[62] A. H. Gnauck, P. J. Winzer, S. Chandrasekhar, X. Liu, B. Zhu, and D. W. Peckham, "10×224-Gb/s WDM transmission of 28-Gbaud PDM 16-QAM on a 50-GHz grid over 1,200 km of fiber," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2010, paper PDPB8. 26

[63] C. R. S. Fludger, J. C. Geyer, T. Duthel, and C. Schulien, "Digital signal processing - from simulation to silicon," in *36th European Conference and Exhibition on Optical Communication (ECOC)*, 2010, paper Tu.5.A.1. 26, 135

[64] J. C. Geyer, A. Bisplinghoff, T. Duthel, C. R. S. Fludger, C. Schulien, and B. Schmauss, "Optimization of the chromatic dispersion equalizer of a 43 Gb/s realtime coherent receiver," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2010, paper OWV8. 26, 135

[65] Z. Wang and J. Crowcroft, "Quality-of-service routing for supporting multimedia applications," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 9, pp. 1228–1234, 1996. 29

[66] P. Mehrotra, G. Karmous-Edwards, and D. Stevenson, "Defining optical plane QoS parameters for OBS networks," in *First International Workshop on Optical Burst Switching (WOBS)*, Oct. 2003. 30

[67] J. He, M. Brandt-Pearce, and S. Subramaniam, "QoS-aware wavelength assignment with BER and latency constraints for all-optical networks," *Journal of Lightwave Technology*, vol. 27, no. 5, pp. 462–474, 2009. 30

[68] K. Yang, S. Ou, K. Guild, and H.-H. Chen, "Convergence of ethernet PON and IEEE 802.16 broadband access networks and its QoS-aware dynamic bandwidth allocation scheme," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 2, pp. 101–116, 2009. 31

[69] S. Azodolmolky, Y. Pointurier, M. Angelou, D. Careglio, J. Solé-Pareta, and I. Tomkos, "A novel impairment aware RWA algorithm with consideration of QoT estimation inaccuracy," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 3, no. 4, pp. 290–299, Apr 2011. 31

[70] W. Vereecken, W. Van Heddeghem, B. Puype, D. Colle, M. Pickavet, and P. Demeester, "Optical networks: How much power do they consume and how can we optimize this?" in *36th European Conference and Exhibition on Optical Communication (ECOC)*, 2010, paper Mo.1.D.1. 31, 88

[71] C. Lange, D. Kosiankowski, R. Hu andlsermann, R. Weidmann, and A. Gladisch, "Energy footprint of telecommunication networks," in *36th European Conference on Optical Communication (ECOC)*, 2010, paper Mo.1.D.2. 31

[72] C. Lange, D. Kosiankowski, R. Weidmann, and A. Gladisch, "Energy consumption of telecommunication networks and related improvement options," *to appear in IEEE Journal of Selected Topics in Quantum Electronics*, 2011. 31, 32, 88

[73] D. C. Kilper, G. Atkinson, S. K. Korotky, S. Goyal, P. Vetter, D. Suvakovic, and O. Blume, "Power trends in communication networks," *to appear in IEEE Journal of Selected Topics in Quantum Electronics*, 2011. 31

[74] GreenTouch, [Online]: http://www.greentouch.org/. 32, 36, 105

[75] R. S. Tucker, "Green optical communications-part I: Energy limitations in transport," *to appear in IEEE Journal of Selected Topics in Quantum Electronics*, 2011. 32

[76] ——, "Green optical communications-part II: Energy limitations in networks," *to appear in IEEE Journal of Selected Topics in Quantum Electronics*, 2011. 32

[77] G. Shen and R. S. Tucker, "Energy-minimized design for IP over WDM networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 1, no. 1, pp. 176–186, 2009. 32, 36, 88, 105, 196

[78] S. Aleksic, "Analysis of power consumption in future high-capacity network nodes," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 1, no. 3, pp. 245–258, 2009. 32

[79] Y. Zhang, P. Chowdhury, M. Tornatore, and B. Mukherjee, "Energy efficiency in telecom optical networks," *IEEE Communications Surveys Tutorials*, vol. 12, no. 4, pp. 441–458, 2010. 32, 35, 36, 196

[80] M. Xia, M. Tornatore, Y. Zhang, P. Chowdhury, C. U. Martel, and B. Mukherjee, "Green provisioning for optical WDM networks," *to appear in IEEE Journal of Selected Topics in Quantum Electronics*, 2011. 35

[81] M. Ruffini, D. O'Mahony, and L. Doyle, "Optical IP switching: A flow-based approach to distributed cross-layer provisioning," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 2, no. 8, pp. 609–624, Aug 2010. 36

[82] C. Chigan, G. W. Atkinson, and R. Nagarajan, "Cost effectiveness of joint multilayer protection in packet-over-optical networks," *Journal of Lightwave Technology*, vol. 21, no. 11, pp. 2694–2704, 2003. 36

[83] N. Vasić and D. Kostić, "Energy-aware traffic engineering," in *1st International Conference on Energy-Efficient Computing and Networking*, ser. e-Energy '10. New York, NY, USA: ACM, 2010, pp. 169–178. 36

[84] J. Chabarek, J. Sommers, P. Barford, C. Estan, and D. Tsiang, "Power awareness in network design and routing," in *IEEE International Conference on Computer Communications (INFOCOM)*, Apr. 2008, pp. 1130–1138. 36

[85] S. J. B. Yoo, "Energy efficiency in the future internet: The role of optical packet switching and optical-label switching," *to appear in IEEE Journal of Selected Topics in Quantum Electronics*, 2011. 37, 63, 196

[86] A. Bianco, D. Camerino, D. Cuda, and F. Neri, "Optics vs. electronics in future high-capacity switches/routers," in *International Conference on High Performance Switching and Routing (HPSR)*, 2009, pp. 1 –6. 38

[87] J. H. Sinsky, "Integration and packaging of devices for 100-Gb/s transmission," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2009, paper OThN6. 38

[88] A. Shacham and K. Bergman, "Building ultralow-latency interconnection networks using photonic integration," *IEEE Micro*, vol. 27, no. 4, pp. 6–20, 2007. 40

[89] A. Shacham, B. G. Lee, and K. Bergman, "A wide-band nonblocking 2×2 switching node for a SPINet network," *IEEE Photonics Technology Letters*, vol. 17, no. 12, pp. 2742–2744, 2005. 40

[90] W. Dally and B. Towles, *Principles and Practices of Interconnection Networks*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003. 41

[91] C. P. Lai, A. Shacham, and K. Bergman, "Demonstration of asynchronous operation of a multiwavelength optical packet-switched fabric," *IEEE Photonics Technology Letters*, vol. 22, no. 16, pp. 1223–1225, 2010. 43, 67, 83

[92] O. Liboiron-Ladouceur, B. Small, and K. Bergman, "Physical layer scalability of WDM optical packet interconnection networks," *Journal of Lightwave Technology*, vol. 24, no. 1, pp. 262–270, 2006. 44

[93] C. P. Lai, M. A. Koroshetz, and K. Bergman, "Improving the soa switching speed of wavelength-striped optical packets using multipulse current injection," in *Photonics in Switching (PS)*, 2010, paper PMC6. 53

[94] M. J. Connelly, *Semiconductor Optical Amplifiers*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2002. 53

[95] J. Bowers, E. Burmeister, and D. Blumenthal, "Optical buffering and switching for optical packet switching," in *Photonics in Switching (PS)*, 2006. 53

[96] R. P. Luijten and R. Grzybowski, "The OSMOSIS optical packet switch for supercomputers," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2009, paper OTuF3. 53, 65

[97] O. Liboiron-Ladouceur, A. Shacham, B. A. Small, B. G. Lee, H. Wang, C. P. Lai, A. Biberman, and K. Bergman, "The data vortex optical packet switched interconnection network," *Journal of Lightwave Technology*, 2008. 53, 82

[98] M. Ikeda, "Switching characteristics of laser diode switch," *IEEE Journal of Quantum Electronics*, vol. 19, no. 2, pp. 157–164, Feb. 1983. 54

[99] S. P. Q. Xu, B. Schmidt and M. Lipson, "Micrometre-scale silicon electro-optic modulator," *Nature*, 2005. 54

[100] N. S. Ribeiro, A. L. Toazza, C. M. Gallep, and E. Conforti, "Rise time and gain fluctuations of an electrooptical amplified switch based on multipulse injection in semiconductor optical amplifiers," *IEEE Photonics Technology Letters*, 2009. 54

[101] C. M. Gallep and E. Conforti, "Reduction of semiconductor optical amplifier switching times by preimpulse step-injected current technique," *IEEE Photonics Technology Letters*, vol. 14, no. 7, pp. 902–904, Jul. 2002. 54

[102] O. Liboiron-Ladouceur and K. Bergman, "Optimization of a switching node for optical multistage interconnection networks," *IEEE Photonics Technology Letters*, vol. 19, no. 20, pp. 1658–1660, Oct. 2007. 58

[103] B. G. Lee, C. P. Lai, J. D. Foster, B. A. Small, and K. Bergman, "Gain uniformity optimization of SOA-based optical packet switching nodes for performance and scalability improvements," *Journal of Optical Networking*, vol. 6, no. 9, pp. 1055–1063, Sep 2007. 61, 62

[104] R. S. Tucker, "The role of optics and electronics in high-capacity routers," *Journal of Lightwave Technology*, vol. 24, no. 12, pp. 4655 –4673, 2006. 63

[105] X. Ye, V. Akella, and S. J. B. Yoo, "Comparative studies of all-optical vs. electrical vs. hybrid switches in datacom and in telecom networks," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2011, paper OThQ4. 63

[106] J. A. Kash, A. F. Benner, F. E. Doany, D. M. Kuchta, B. G. Lee, P. K. Pepeljugoski, L. Schares, C. L. Schow, and M. Taubenblatt, "Optical interconnects in exascale supercomputers," in *23rd Annual Meeting of the IEEE Photonics Society*, 2010, paper WR1. 65

[107] S. H. Chin, A. Franzan, D. K. Hunter, and I. Andanovic, "Synchronisation schemes for optical networks," *IEE Proceedings Optoelectronics*, vol. 147, no. 6, pp. 423–427, Dec. 2000. 66

[108] D. Dai and D. K. Panda, "How can we design better networks for dsm systems?" in *Proceedings of the Second International Workshop on Parallel Computer Routing and Communication*. London, UK: Springer-Verlag, 1997, pp. 171–184. 66

[109] L. Tancevski, S. Yegnanarayanan, G. Castanon, L. Tamil, F. Masetti, and T. McDermott, "Optical routing of asynchronous, variable length packets," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 10, pp. 2084–2093, Oct. 2000. 66

[110] J. P. Mack, K. N. Nguyen, J. M. Garcia, E. F. Burmeister, M. M. Dummer, H. N. Poulsen, B. Stamenic, G. Kurczveil, K. Hollar, L. A. Coldren, J. E. Bowers, and D. J. Blumenthal, "Asynchronous 2×2 optical packet synchronization, buffering, and forwarding," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2010, p. OThN1. 66

[111] R. Nejabati, G. Zervas, D. Simeonidou, M. J. O'Mahony, and D. Klonidis, "The "OPORON" project: Demonstration of a fully functional end-to-end asynchronous optical packet-switched network," *Journal of Lightwave Technology*, vol. 25, no. 11, pp. 3495–3510, 2007. 67

[112] A. Shacham, C. P. Lai, and K. Bergman, "Experimental demonstration of an optical interconnection network with asynchronous transmission," in *33th European Conference and Exhibition on Optical Communication (ECOC)*, 2007, paper 6.5.6. 67

[113] C. P. Lai and K. Bergman, "Implementing an optical qos encoding scheme in an optical packet switching fabric test-bed," *IEEE Photonics Technology Letters*, vol. 22, no. 20, pp. 1518–1520, 2010. 74

[114] C. P. Lai, H. Wang, A. Shacham, and K. Bergman, "Priority encoding scheme for contention resolution in optical packet-switched networks," in *34th European Conference and Exhibition on Optical Communication (ECOC)*, 2008, paper P.5.7. 74, 75

[115] C. P. Lai, F. Fidler, and K. Bergman, "Experimental demonstration of QoS-aware cross-layer packet protection switching," in *35th European Conference on Optical Communication (ECOC)*, 2009, paper 2.5.3. 74, 121, 135, 137, 165, 171, 172, 179

[116] S. Yao, S. J. Ben Yoo, and B. Mukherjee, "A comparison study between slotted and unslotted all-optical packet-switched network with priority-based routing," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, vol. 2, Mar. 2001, paper TuK2-1. 74

[117] A. G. P. Rahbar and O. Yang, "Prioritized retransmission in slotted all-optical packet-switched networks," *Journal of Optical Networking*, vol. 5, no. 12, pp. 1056–1070, Dec 2006. 74

[118] B. G. Bathula and V. M. Vokkarane, "Qos-based manycasting over optical burst-switched (obs) networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 1, pp. 271–283, 2010. 75, 159

[119] C. P. Lai and K. Bergman, "Broadband multicasting for wavelength-striped optical packets," *Submitted to Journal of Lightwave Technology*, 2011. 87

[120] G. N. Rouskas, "Optical layer multicast: rationale, building blocks, and challenges," *IEEE Network*, vol. 17, no. 1, pp. 60–65, 2003. 87

[121] M. A. Marsan, A. Bianco, P. Giaccone, E. Leonardi, and F. Neri, "Multicast traffic in input-queued switches: optimal scheduling and maximum throughput," *IEEE/ACM Transactions on Networking*, vol. 11, no. 3, pp. 465–477, 2003. 88

[122] M. Andrews, S. Khanna, and K. Kumaran, "Integrated scheduling of unicast and multicast traffic in an input-queued switch," in *18th IEEE International Conference on Computer Communications (INFOCOM)*, vol. 3, Mar. 1999, pp. 1144–1151. 88, 113

[123] N. Yan, I. T. Monroy, H.-D. Jung, T. Koonen, A. Teixeira, and T. Silveira, "Optical multicast technologies by multi-wavelength conversion for optical routers," in *International Conference on Communication Technology (ICCT)*, 2006, pp. 1–4. 88

[124] Q. Huang and W.-D. Zhong, "Wavelength-routed optical multicast packet switch with improved performance," *Journal of Lightwave Technology*, 2009. 88

[125] D. Tutsch and G. Hommel, "Performance of buffered multistage interconnection networks in case of packet multicasting," in *Proceedings Advances in Parallel and Distributed Computing*, Mar. 1997, pp. 50–57. 88

[126] X. Liu, H. Wang, and Y. Ji, "Hybrid multicast mode in all-optical networks," *IEEE Photonics Technology Letters*, vol. 19, no. 16, pp. 1212–1214, 2007. 88, 171, 180

[127] W. S. Hu and Q. J. Zeng, "Multicasting optical cross connects employing splitter-and-delivery switch," *IEEE Photonics Technology Letters*, vol. 10, no. 7, pp. 970 –972, Jul. 1998. 90

[128] C. P. Lai and K. Bergman, "Demonstration of programmable broadband packet multicasting in an optical switching fabric test-bed," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2009, paper OTuA5. 90, 92, 137, 176

[129] C. P. Lai and K. Bergman, "Network architecture and test-bed demonstration of wavelength-striped packet multicasting," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2010, paper OWI4. 92, 105

[130] S. Okamoto, A. Watanabe, and K.-I. Sato, "Optical path cross-connect node architectures for photonic transport network," *Journal ofLightwave Technology*, vol. 14, no. 6, pp. 1410–1422, Jun. 1996. 93

[131] L. Chen, S. H. Low, M. Chiang, and J. C. Doyle, "Cross-layer congestion control, routing and scheduling design in ad hoc wireless networks," in *25th IEEE International Conference on Computer Communications (INFOCOM)*, 2006, pp. 1–13. 120

[132] M. Chiang, S. H. Low, and A. R. Calderbank, "Layering as a optimization decomposition: A mathematical theory of network architectures," *Proc. of IEEE*, vol. 95, no. 1, pp. 255–312, 2007. 120

[133] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1452–1463, Aug. 2006. 120

[134] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Communications Magazine*, vol. 41, no. 10, pp. 74–80, oct 2003. 120

[135] E. Salvadori, Y. Ye, C. Saradhi, A. Zanardi, H. Woesner, M. Carcagni, G. Galimberti, G. Martinelli, A. Tanzi, and D. La Fauci, "Distributed optical control plane architectures for handling transmission impairments in transparent optical networks," *Journal of Lightwave Technology*, 2009. 120

[136] C. P. Lai, H. Wang, and K. Bergman, "Cross-layer communication with an optical packet switched network via a message injection control interface," *IEEE Photonics Technology Letters*, vol. 20, no. 12, pp. 967–969, 2008. 123

[137] M. Enachescu, Y. Ganjali, A. Goel, N. McKeown, and T. Roughgarden, "Routers with very small buffers," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2006, pp. 1–11. 123

[138] B. A. Small, A. Shacham, and K. Bergman, "A modular, scalable, extensible, and transparent optical packet buffer," *Journal of Lightwave Technology*, vol. 25, no. 4, pp. 978–985, 2007. 124

[139] A. Shacham, B. A. Small, and K. Bergman, "A novel optical buffer architecture for optical packet switching routers," in *32th European Conference and Exhibition on Optical Communication (ECOC)*, 2006, paper We1.4.4. 124

[140] A. Shacham and K. Bergman, "Optical packet buffers with active queue management," in *11th Conference on Optical Network Design and Modeling (ONDM)*, 2007, paper Th2.1. 124

[141] H. Wang, C. P. Lai, A. Shacham, and K. Bergman, "Experimental demonstration of network congestion control with a programmable optical packet injection buffer," in *20th Annual Meeting of the IEEE Lasers and Electro-Optics Society (LEOS)*, 2007, paper ThG2. 124

[142] C. P. Lai, H. Wang, and K. Bergman, "Interface optical buffer and packet-switched network cross-layer signaling demonstration," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2008, paper OThI5. 124

[143] O. Gerstel, I. Leung, G. Nicholl, H. Sohel, W. Wakim, and K. Wollenweber, "Near-hitless protection in ipodwdm networks," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2008, paper NWD4. 133, 136, 156, 171

[144] T. Mizuochi, "Next generation FEC for optical communication," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2008, paper OTuE5. 134, 136

[145] ns-2 Wiki: The Network Simulator-ns-2, [Online]: http://nsnam.isi.edu/nsnam/index.php/. 136, 146

[146] F. Fidler, C. P. Lai, and K. Bergman., "Geni embedded real-time measurements, technical note 3, discrete-event network simulations," GENI, Tech. Rep., 2009, [Online]: http://groups.geni.net/. 136, 149

[147] F. Fidler, P. J. Winzer, C. P. Lai, M. K. Thottan, and K. Bergman, "Cross-layer simulations of fast packet protection mechanisms," in *35th European Conference and Exhibition on Optical Communication (ECOC)*, 2009, paper 2.5.2. 136

[148] H. Bulow, W. Baumert, H. Schmuck, F. Mohr, T. Schulz, F. Kuppers, and W. Weiershausen, "Measurement of the maximum speed of pmd fluctuation in installed field fiber," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, vol. 2, Feb. 1999, pp. 83–85. 147

[149] Y. Sun, A. Srivastava, J. L. Zyskind, J. W. Sulhoff, C. Wolf, and R. W. Tkach, "Fast power transients in wdm optical networks with cascaded edfas," *Electronics Letters*. 147, 148

[150] S. Chandrasekhar and X. Liu, "Impact of channel plan and dispersion map on hybrid dwdm transmission of 42.7-Gb/s DQPSK and 10.7-Gb/s OOK on 50-GHz Grid," *IEEE Photonics Technology Letters*, vol. 19, no. 22, pp. 1801–1803, 2007. 147

[151] P. J. Winzer, M. Pfennigbauer, and R.-J. Essiambre, "Coherent crosstalk in ultradense wdm systems," *Journal of Lightwave Technology*, vol. 23, no. 4, pp. 1734–1744, 2005. 151

[152] C. P. Lai, B. G. Bathula, V. M. Vokkarane, and K. Bergman, "QoS-aware cross-layer multicasting for optical packet-switched networks: Simulation exploration and test-bed demonstration," in *36th European Conference and Exhibition onOptical Communication (ECOC)*, 2010, paper Tu.3.D.2. 157

[153] A. E. Willner, "The optical network of the future: Can optical performance monitoring enable automated, intelligent and robust systems?" *Optics Photonics News*, vol. 17, no. 3, pp. 30–35, 2006. 169

[154] J. J. Martinez, M. I. Roche, A. Villafranca, C. D. Heras, J. M. Subias, J. Pelayo, I. Garces, P. Blasco, and E. Pellejer, "Novel in-band osnr monitoring system based on a polarization interferometer,," in *36th European Conference and Exhibition on Optical Communication (ECOC)*, 2010, paper P3.18. 169

[155] J.-Y. Yang, M. R. Chitgarha, L. Zhang, and A. E. Willner, "PMD and OSNR insensitive 40-Gb/s OOK/DPSK chromatic dispersion monitoring using a delay-line interferometer and a $<$ 10-ghz photodetector," in *Conference on Lasers and Electro-Optics (CLEO) and Quantum Electronics and Laser Science Conference (QELS)*, May 2010, paper JThE50. 169

[156] V. Ribeiro, L. Costa, A. Teixeira, R. Nogueira, and M. Lima, "Chromatic-dispersion-monitoring scheme using a Mach-Zehnder Interferometer and Q-Factor Calculation," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 2, no. 1, pp. 10–19, Jan 2010. 169, 185

[157] S. M. R. M. Nezam, Y.-W. Song, C. Yu, J. E. McGeehan, A. B. Sahin, and A. E. Willner, "First-order PMD monitoring for NRZ data using RF clock regeneration techniques," *Journal of Lightwave Technology*, vol. 22, no. 4, pp. 1086–1093, 2004. 169

[158] I. Tomkos, D. Klonidis, and S. Azodolmolky, "Impairment aware networking and relevant resiliency issues in all-optical networks," in *34th European Conference and Exhibition on Optical Communication (ECOC)*, 2008, paper We3D1. 169

[159] T. Anderson, J. C. Li, D. Hewitt, and O. Jerphagnon, "Optical performance monitoring for intelligent networks," in *35th European Conference and Exhibition on Optical Communication (ECOC)*, 2009, paper 9.3.3. 170

[160] X. Yi, W. Chen, and W. Shieh, "An OSNR monitor for optical packet switched networks," *IEEE Photonics Technology Letters*, vol. 18, no. 13, pp. 1448–1450, 2006. 170, 171

[161] X. Liu, Y.-H. Kao, S. Chandrasekhar, I. Kang, S. Cabot, and L. L. Buhl, "Osnr monitoring method for OOK and DPSK based on optical delay interferometer," *IEEE Photonics Technology Letters*, vol. 19, no. 15, pp. 1172–1174, 2007. 170

[162] R. Vilar, J. Garcia, G. Tremblay, Y. Kim, S. Larochelle, F. Ramos, and J. Marti, "Osnr monitoring at high-speeds using a fbg-based correlator in optical packet-switched networks," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2009, paper JWA8. 170

[163] R. Vilar, J. Garcia, Y. Kim, S. LaRochelle, R. Llorente, and F. Ramos, "Path monitoring for restoration functions in optical packet-switched networks," in *11th International Conference on Transparent Optical Networks (ICTON)*, 2009, paper We.B2.6. 171

[164] S. Gupta and B. Jalali, "Time stretch enhanced recording oscilloscope," *Applied Physics Letters*, vol. 94, no. 4, pp. 041 105–041 105–3, Jan. 2009. 185, 187, 189

[165] C. P. Lai, A. Fard, B. Buckley, B. Jalali, and K. Bergman, "Cross-layer signal monitoring in an optical packet-switching test-bed via real-time burst sampling," in *23rd Annual Meeting of the IEEE Photonics Society*, 2010, paper ME3. 185

[166] I. Shake, H. Takara, and S. Kawanishi, "Simple Q factor monitoring for BER estimation using opened eye diagrams captured by high-speed asynchronous electrooptical sampling," *IEEE Photonics Technology Letters*, vol. 15, no. 4, pp. 620–622, 2003. 185

[167] A. Motafakker-Fard, S. Gupta, and B. Jalali, "Digital equalization of ultrafast data using real-time burst sampling," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2010, paper OTuN3. 186

[168] B. W. Buckley, A. Fard, and B. Jalali, "Time-stretch analog-to-digital conversion using phase modulation and broadband balanced coherent detection for improving resolution," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2011, paper OThW4. 186, 212

[169] A. Fard, J.-Y. Yang, B. Buckley, J. Wang, M. Chitgarha, L. Zhang, A. E. Willner, and B. Jalali, "100-Gb/s RZ-DQPSK signal monitoring using time-stretch enhanced recording oscilloscope," in *Conference on Lasers and Electro-Optics (CLEO) and Quantum Electronics and Laser Science Conference (QELS)*, May 2011, paper CFP1. 187

[170] A. L. Chiu and J. Strand, "An agile optical layer restoration method for router failures," *IEEE Network*, vol. 17, no. 2, pp. 38–42, 2003. 196

[171] C. P. Lai, D. Brunina, C. Ware, B. G. Bathula, and K. Bergman, "Demonstration of cross-layer failure recovery for reconfigurable optical switching fabrics," *Submitted to IEEE Photonics Technology Letters*, 2011. 197, 217

[172] C. P. Lai, C. Ware, B. G. Bathula, D. Brunina, and K. Bergman, "Intelligent highly-functional cross-layer optimized interfaces for future access/aggregation networks," *To be presented at 13th International Conference on Transparent Optical Networks (ICTON)*, 2011. 209

[173] C. Ware, B. G. Bathula, C. P. Lai, D. Brunina, and K. Bergman, "Cross-layer reconfigurable optical network: fast failure recovery in testbed for routing algorithms," *To be presented at 13th International Conference on Transparent Optical Networks (ICTON)*, 2011. 209

[174] H. Wang, A. S. Garg, K. Bergman, and M. Glick, "Design and demonstration of an all-optical hybrid packet and circuit switched network platform for next generation data centers," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2010, paper OTuP3. 211

[175] W. Zhang, A. S. Garg, H. Wang, C. P. Lai, J. Wu, J. Lin, and K. Bergman, "Experimental demonstration of 10 gigabit ethernet-based optical interconnection network interface for large-scale computing systems," *Submitted to European Conference and Exhibition on Optical Communication (ECOC)*, 2011. 229

[176] FFmpeg, [Online]: http://www.ffmpeg.org/. 232

[177] NetFPGA, [Online]: http://netfpga.org/. 234

[178] RENCI - BEN: Breakable Experimental Network, [Online]: https://ben.renci.org/. 234

[179] "The Intelligent Path To A Service-Enabling Infrastructure [White Paper]," Ciena, Tech. Rep., 2011, [Online]: www.ciena.com/. 238

[180] D. Verchere, "Cloud computing over telecom network," in *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC)*, 2011, paper OMW1. 240

[181] Columbia University WiMAX Campus Deployment and Installation, [Online]: http://groups.geni.net/geni/wiki/WIMXCOLUM/CUWIMAX. 243